nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

NEAR HORIZON

A warm terrestrial planet in orbit around Proxima Centauri, closest star to the Sun PAGES 408 & 437



PHARMACEUTICALS

VITAL DRUGS MADE CHEAPER

Non-profit initiative targets neglected diseases

PAGE 388

SUSTAINABILITY

CITIES COOK CARBON BOOKS

Overhaul urban emissions accounts

PAGE 391

ROBOTICS

THE SOFT MACHINE

Self-powered 'octobot' breaks new ground

PAGES 400 & 451

NATURE.COM/NATURE

25 August 2016 £10

Vol. 536, No. 7617

9 770028 083095 34

THIS WEEK

EDITORIALS

WORLD VIEW China must back up its plan for soil quality with legislation **p.375**

DISCO FEVER Fluorescent rodents light up organ analyses **p.376**

NORTH FACE Alaskan skull fossil is a new species of river dolphin **p.377**



Go forth and replicate!

To make replication studies more useful, researchers must make more of them, funders must encourage them and journals must publish them.

No scientist wants to be the first to try to replicate another's promising study: much better to know what happened when others tried it. Long before replication or reproducibility became major talking points, scientists had strategies to get the word out. Gossip was one. Researchers would compare notes at conferences, and a patchy network would be warned about whether a study was worth building on. Or a vague comment might be buried in a related publication. Tell-tale sentences would start "In our hands", "It is unclear why our results differed …" or "Interestingly, our results did not …".

What might seem obvious — a paper on attempts and outcomes — was almost never an option. Many journals refused to consider replication studies, and a lot of researchers had no desire to start a feud if their results did not match. So scientists not in the know might waste time exploring a blind alley or be wary about truly promising research.

Things are improving. Nowadays, researchers who want to tell the scientific community about their replication studies have multiple ways to do so. They can chronicle their attempts on a blog, post on a preprint server or publish peer-reviewed work in those journals that do not require novelty. Just this year, the online platform F1000 launched the dedicated Preclinical Reproducibility and Robustness channel for refutations, confirmations or more nuanced replication studies. Other titles, including *Scientific Data* and the *American Journal of Gastroenterology*, have openly solicited replication attempts and negative results. In 2013, after controversial work on whether bioactive RNA molecules could cross from the digestive tract to the bloodstream, *Nature Biotechnology* declared itself "receptive to replication", provided that such studies illuminate crucial research questions (*Nature Biotechnol.* **31,** 943; 2013).

The psychology community is a leader in this: *Perspectives on Psychological Science* has begun publishing a new type of article, and pioneering a new form of collaboration. It asks psychologists to nominate an influential study for replication and to draw up a plan. The original author is invited to offer suggestions on the protocol, multiple labs volunteer to collect data, and results — whatever they may be — are published as a registered replication report (RRR). So far, three have been published, each with a perspective by the original authors.

GET IT OUT THERE

Yet it would be inefficient to pursue such projects for more than a sliver of publications. Most replication attempts are not organized collaborations, but individual laboratories testing the next stage of their research. If those results were shared, science would benefit.

Why doesn't this happen more often? Because the replication ecosystem, such as it is, lacks visibility, value and conventions.

When a researcher happens on an exciting paper, there is no easy way to learn about replication attempts. Replication studies are not automatically or consistently linked to original papers on journal websites, PubPeer or PubMed. When a replication attempt is mentioned in passing in a broader study, there is no way to capture it. Journals cannot be expected to curate all replication attempts of papers they publish, although they should support technology that aggregates and disseminates that information. And they should be open to publishing in-depth replication attempts for original papers. For example, *Scientific Reports* encourages critique by offering to waive its article-processing charge for a peer-reviewed refutation of an article published in the journal.

Increased visibility would raise the value of a replication attempt, but also increase the risk of retaliation against replicators. There is little reward for taking that risk. A published replication currently does little to raise the esteem of the replicator with hiring committees or grant reviewers. This creates a chicken–egg problem — researchers don't want to conduct and publish rigorous replication studies because they are not valued, and replication studies are not valued because few are published. Commendably, funders such as the Laura and John Arnold Foundation in the United States and the Netherlands Organisation for Scientific Research are explicitly supporting replication studies, and setting high expectations for publication. Scientists can help to ensure that such studies are valued by citing them and by discussing them on social media.

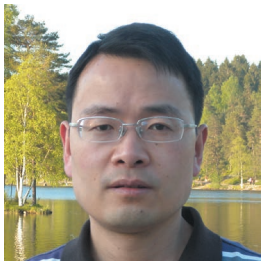
"Conventions around replication are in their infancy — even the vocabulary is inadequate."

Conventions around replication studies are in their infancy — even the vocabulary is inadequate. Editors who coordinate RRRs strive to avoid loaded labels such as 'successful' and 'failed' replications. The Reproducibility Initiative, a project to help labs coordinate independent replications of their own work, also shied away from similar pronouncements after its first study. A paper is a jumble of context, experiments, results, analysis and informed speculation. Outcomes can depend on apparently trivial differences in methods, such as how vigorously reagents are mixed, as one collaboration painstakingly discovered (W. C. Hines *et al. Cell Rep.* **6,** 779–781; 2014).

Neither are there conventions for interactions between replicators and the original authors. Some original authors have refused to share data or methodological details. In other cases, some replicators broadcast their attempts without first trying to resolve inconsistencies, a practice that leaves them more open to charges of incompetence. (Thankfully, both replicators and original authors are now backing away from name-calling.) As replication becomes more mainstream, we trust that the community will establish reasonable standards of conduct.

To foster better behaviour, replication attempts must become more common. We urge researchers to open their file drawers. We urge authors to cooperate with reasonable requests for primary data, to assume good intent and to write papers — and keep records — assuming that others will want to replicate their work. We urge funders and publishers to support tools that help researchers to thread the literature together. We welcome, and will be glad to help disseminate, results that explore the validity of key publications, including our own. ■

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.


XUEYAN HU

China's soil plan needs strong support

The government must accompany its action plan on soil quality with effective laws and remediation measures, says **Hong Yang**.

The impact of the booming Chinese economy on the quality of the nation's air and water has garnered a lot of attention recently. Now, focus is turning to another polluted realm: the very ground beneath China's feet.

Chinese researchers are still digesting the implications of a national action plan on soil quality released in May. But two problems are already clear, neither of them unique to China. The first is that remediation of polluted soil is very expensive. The second is that binding legislation must be in place to make sure the necessary money is spent.

The consequences of failing to act properly were graphically demonstrated in April. Nearly 500 teenagers at a school constructed near three former chemical plants in Changzhou, east China, were found to have developed health problems ranging from nosebleeds to lymphoma and leukaemia. Preliminary investigations found the soils and groundwater to be laced with heavy metals and organic pollutants. In particular, chlorobenzene was found at concentrations nearly 10,000 times the safe level.

Ironically, problems like this are the consequence of concerns about pollution. Dirty factories and industrial plants moved from urban areas to the suburbs and rural locations. Their former sites were demolished, and schools and houses were built in their place.

There are thousands of brownfield sites across China, and many are heavily polluted. Serious pollution risks were confirmed by the authorities in China's soil-pollution survey, once guarded as a national secret but publicly released in April 2014. It claimed that 16% of China's soils exceed national standards for pollution with heavy metals and pesticides and, remarkably, that 34.9% of brownfields exceeded national pollution standards.

But a lack of transparent detailed information on locations and pollution levels at specific sites hinders public awareness and efforts to tackle the problem. Unlike the often visible pollution of air or water, soil contamination is difficult to detect without professional equipment. Many brownfield sites have already been used for businesses or housing with insufficient remediation. And until something changes, seriously polluted sites will continue to imperil the environment.

The nationwide action plan calls for new laws to monitor, prevent and remediate the soil pollution. It is very ambitious, aiming to curb new soil pollution by 2020 and incrementally improve soil quality across the country by the middle of this century.

To implement the plan fully would be extremely costly, at more than 7 trillion yuan (US$1.06 trillion) according to one estimate. To put that into perspective, the Chinese government's investment in soil remediation this year is around 9 billion yuan.

How can the gap be closed?

The common 'polluter pays' principle applies in China, but soil pollution can take a long time to emerge and be detected. That makes it difficult to track and locate responsible parties, and these historical polluters often have no means to pay. Other funding mechanisms have to be considered, for example from the World Bank and Global Environment Facility. Land developers also need to meet more of the clean-up cost before they are granted the rights to use and sell brownfield sites. Even then, the costs will probably be too high to remediate properly, so China must consider a range of other funding instruments, from environmental taxes and clean-up subsidies to loan guarantees and insurance.

The new plan does not specify which technologies should be used for soil remediation, and this is a problem. At present, many projects try to shorten the treatment period and reduce costs by removing just the polluted topsoil, and so the contamination remains in deeper ground and water. The removed topsoil is landfilled, which just moves the problem somewhere else, or incinerated, which releases contaminants into the air. Soil-washing is an option, but it generates secondary wastes that require additional decontamination. Improper decontamination measures can even make a site worse by releasing buried pollutants. This seems to be what happened at the Changzhou school.

LOW-COST AND **EFFECTIVE** NEW TECHNOLOGIES ARE URGENTLY **NEEDED.**

Low-cost and effective new technologies are urgently needed. Plant-based bioremediation approaches have been developed to treat arsenic and cadmium pollution in China. More research on other biological approaches (use of earthworms, nematodes and phytoremediation) should be launched.

In common with previous environmental efforts in China, success of the soil action plan will depend on the strength of the law that supports it. Legislation in this area is outdated, and a replacement has been promised by 2020. It should include firm criteria and measures to determine the effectiveness of remediation. (At present, some areas in China use guidelines prepared by the US Environmental Protection Agency, which often do not properly apply to local circumstances).

China is not alone in facing risks caused by soil pollution. Similar major problems have occurred in other developing countries undergoing rapid urbanization, including India and Brazil. As a world laboratory for new technology to decontaminate brownfields, China's approach to managing the problem could benefit other countries. By 2050, another 2.5 billion people are expected to be living in cities. The brownfields must be treated effectively to give these new urban residents clean land. ■

Hong Yang *is a researcher at the Norwegian Institute of Bioeconomy Research (NIBIO) and the University of Oslo, Norway.*
e-mail: hongyanghy@gmail.com

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

RESEARCH HIGHLIGHTS

Selections from the scientific literature

ECOLOGY

Longing for home undid cave bears

When looking for a spot to hibernate, ancient cave bears stuck with family.

Cave bears (*Ursus spelaeus*) went extinct 24,000 years ago, whereas the related brown bears (*Ursus arctos*) still thrive. To explain these differing fates, a team led by Gloria Fortes and Axel Barlow at the University of Potsdam, Germany, obtained mitochondrial genomes of 31 cave bears and 15 ancient brown bears from caves in Spain. Cave bears from the same caves tended to share mitochondrial DNA, which is inherited maternally, whereas brown-bear caves contained multiple maternal lines.

These patterns suggest that cave bears nearly always hibernated in their native caves, which may have contributed to their demise in the face of competition from other bears, the scientists say.

Mol. Ecol. **http://doi.org/bpgr (2016)**

ECONOMICS

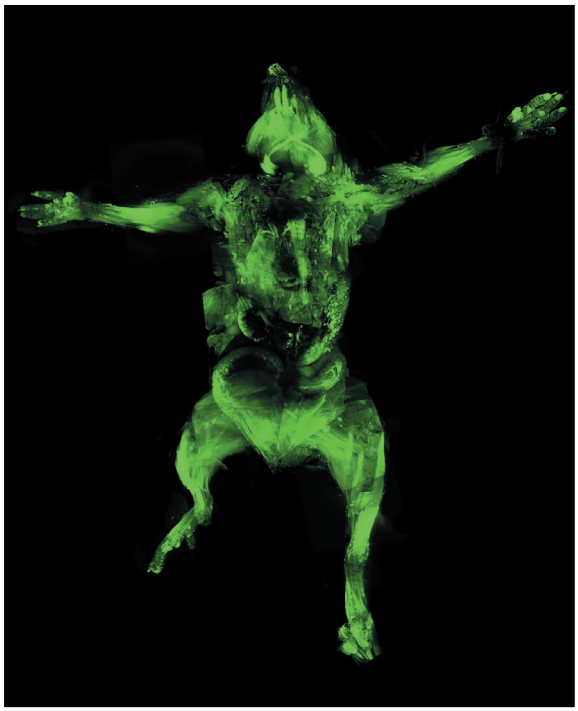
Satellites map world poverty

Computer scientists have used satellite imagery and machine-learning techniques to make detailed maps of regions where poverty is common.

Neal Jean and his colleagues at Stanford University in California focused on Nigeria, Tanzania, Uganda, Malawi and Rwanda and combined various data sets, including daytime images that identify features such as paved roads and metal roofs, to estimate local household consumption and income. When identifying areas where incomes are below the international poverty line of US$1.90 per person per day, the team's algorithm outperforms night-light maps (an alternative but limited indicator of economic activity). It also probes hard-to-reach areas, including regions not accessed by household surveys — such as those conducted by the World Bank — which are costly and infrequently conducted.

The method could prove useful for targeting social programmes and determining when and where they fail.

Science 353, **790–794 (2016)**



CELL BIOLOGY

See-through rodents

An imaging technique lets scientists peer through the skin of a whole mouse or rat to examine its organs after death.

Ali Ertürk of the Ludwig Maximilians University of Munich in Germany and his colleagues created a technique called ultimate DISCO (uDISCO), which removes pigments and lipids from the tissues of dead animals using an organic solvent. This leaves the organs and skin intact but transparent, while preserving genetically encoded fluorescent proteins. The method revealed the nervous system of a mouse in stark detail.

uDISCO also shrinks bodies by up to 65%, making it possible to image whole animals using light-sheet microscopy, which excels at imaging smaller samples.

Nature Methods **http://dx.doi.org/10.1038/nmeth.3964 (2016)**

GENOMICS

Medicine less precise for some

A lack of ethnic diversity in people whose genomes have been sequenced is complicating precision medicine for people with non-European ancestry.

David Goldstein of Columbia University in New York City and Slavé Petrovski of the Royal Melbourne Hospital in Australia examined data from the Exome Aggregation Consortium (ExAC), which contains sequences from 60,252 people, 60.9% of whom have European ancestry. When they compared genetic variants from a cohort of 5,094 people with variants found in the ExAC and other data sets, the comparisons yielded a shorter list of potentially causative variants in people with European ancestry (6.6, on average) than in people with non-European ancestry (9.9–12.7 candidate variants, depending on ethnicity).

Precision medicine is much more precise for people with European than non-European ancestry owing to undersampling of non-European populations, the authors write.

Genome Biol. **http://doi.org/bphp (2016)**

PARTICLE PHYSICS

Neutrino search closes in

Scientists are getting closer to discovering whether neutrinos and antineutrinos are in fact the same particles — known as Majorana neutrinos.

The theory, proposed by Italian physicist Ettore Majorana in the 1930s, could

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

BOERSMA AND PYENSON/CC-BY 4.0

explain why neutrinos have mass and why the Universe contains more matter than antimatter. Azusa Gando at Tohoku University in Sendai, Japan, and her colleagues in the KamLAND-Zen Collaboration carried out the most sensitive search so far for radioactive decay indicative of Majorana neutrinos, using an underground detection facility containing a huge balloon filled with purified xenon.

The team's results, although negative, constrain the upper limit of the mass of Majorana neutrinos to 61–165 millielectronvolts. However, the detector's sensitivity must be pushed even further to prove Majorana's theory, the researchers say.

Phys. Rev. Lett. 117, **082503 (2016)**

MATERIALS SCIENCE

Bulk production of mother-of-pearl

Artificial mother-of-pearl can be made by mimicking the natural process of mineralization.

Mother-of-pearl, or nacre, is remarkably strong yet biodegradable. However, its complex layered structure, in which mineral plates form in an organic scaffold, makes it difficult to recreate in bulk. Shu-Hong Yu at the University of Science and Technology of China in Hefei and his colleagues built their own matrix by growing sheets of ice, which squeezed a solution of the biopolymer chitosan into solid layers. They then pumped this scaffold with materials to grow calcium carbonate, and pressed the stack to form synthetic nacre.

The synthetic version has similar mechanical properties to its natural counterpart and takes just two weeks to grow. This method could be used to produce materials for use in the aerospace industry or as armour, say the authors.

Science **http://doi.org/bpk2 (2016)**

ASTRONOMY

Dark-matter evidence weakens

A survey of X-ray light from galaxy clusters has found no evidence of dark matter decaying, in the latest in a series of contradictory results.

In 2014, two separate teams found an unexpected bump in the energy spectra of dozens of galaxy clusters. Emissions at 3.55 kiloelectronvolts (keV) were seen as a possible sign of the decay of 'sterile' neutrinos with a mass of 7.1 keV. Physicists have hypothesized that these heavier cousins of the three known types of neutrino are possible components of dark matter.

Now astronomer Florian Hofmann at the Max Planck Institute for Extraterrestrial Physics in Garching, Germany, and his team have analysed publicly available data from the Chandra X-ray Observatory concerning 33 galaxy clusters (11 of which were not included in the original studies). Their search found no evidence of an anomalous bump at around 3.55 keV.

Astron. Astrophys. 592, **A112 (2016)**

ANTHROPOLOGY

Early humans were picky dressers

Ancient clothing is rarely preserved, but two independent teams have discovered what early humans wore to cope with the cold European weather.

Mark Collard at Simon Fraser University in Burnaby, Canada, and his colleagues compared the animals that modern indigenous groups used to make cold-weather clothing with the bone types found at early human and Neanderthal sites. Remains from animals with fur, such as foxes and rabbits, were more common at early-human sites, whereas bones from deer, bovids and several other animals were found at both types of site equally. This suggests that early humans used fur to sew specialized cold-weather apparel, but that Neanderthals relied on simpler animal-skin capes, the authors say.

In a separate paper, Niall O'Sullivan at the Institute for Mummies and the Iceman in Bolzano, Italy, and his team sequenced mitochondrial DNA from garments worn by Ötzi, the 5,300-year-old ice mummy. His coat, leggings (**pictured** left) and loincloth were made from the skins of domestic cattle, sheep and goats, whereas his hat and quiver (pictured right) used brown-bear fur and roe-deer skin.

J. Anthropol. Archaeol. **http://doi.org/bn82 (2016);** ***Sci. Rep.*** 6, **31279 (2016)**



INSTITUTE FOR MUMMIES AND THE ICEMAN

EVOLUTION

New dolphin species found

A fossilized dolphin skull in the Smithsonian collection has been identified as an entirely new species 65 years after it was dug out of the ground in Alaska.

Alexandra Boersma and Nicholas Pyenson of the Smithsonian Institution's National Museum of Natural History in Washington DC identified a 23-centimetre-long skull (**pictured**) as a new genus and species in a family called the Allodelphinidae. The extinct animal is closely related to today's South Asian river dolphin (*Platanista gangetica*). The fossil dates from around 25 million years ago, a few million years after cetaceans diverged into toothed whales and filter-feeding baleen whales, and is one of the earliest examples found of the former group.



It is also the most northern specimen yet discovered of the Allodelphinidae, and the researchers have dubbed the animal *Arktocara yakataga* — the face of the north.

PeerJ 4, **e2321 (2016)**

CELL BIOLOGY

Stem cells predict drug safety

Heart muscle cells derived from individual patients' stem cells could be used to test the safety of a drug before it's administered — a boon for precision medicine.

Elena Matsa and Joseph Wu of Stanford University in California and their colleagues made heart muscle cells from induced pluripotent stem cells derived from seven people. They then exposed the muscle cells to one of two drugs that have been linked to heart problems in some people: rosiglitazone and tacrolimus. The results showed differences in how the cells responded to the drugs: one cell line, for example, showed signs of increased stress after treatment that were not observed in cells from other patients.

The approach could one day be used to tailor treatment regimens to individuals and to test drug candidates for potential toxicity before they enter clinical trials.

Cell Stem Cell **http://dx.doi.org/10.1016/j.stem.2016.07.006 (2016)**

NATURE.COM

For the latest research published by *Nature* visit:
www.nature.com/latestresearch

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

SEVEN DAYS *The news in brief*

HEALTH

Poor diet plan

British health campaigners have pulled to pieces a UK government plan to tackle childhood obesity. The plan includes a levy on high-sugar soft drinks. But the final version, unveiled on 18 August, is significantly weaker than some researchers had hoped. Among the critics is the British Medical Association, which said that the government had "rowed back" on promises to crack down on the problem and instead produced a "weak plan" with "pointless" voluntary targets.

Feverish campaign

A logistically challenging emergency vaccination campaign, launched last week, aims to stop the spread of deadly yellow fever in Angola and the Democratic Republic of the Congo. Working with the countries' health ministries and 56 global partners, the World Health Organization is coordinating the vaccination of 14 million people in more than 8,000 locations, including urban areas and hard-to-reach border regions. Since December, more than 400 people in the region have died in the worst outbreak of the mosquito-borne disease for 30 years.

NUMBER CRUNCH

23%

The share of 'bee-friendly' garden plants purchased in the United States that contain neonicotinoid pesticides, which are highly toxic to bees. Until two years ago, more than half of garden plants sold by large US retailers were pretreated with the bee-toxic insecticides.

Source: Friends of the Earth



ANDREW BURTON/GETTY

Villagers vote to leave for good

People in the Alaskan village of Shishmaref voted on 16 August to abandon their tiny island northeast of the Bering Strait and move onto the mainland because of erosion due to global warming and rising sea level. The Inuit village, containing around 600 people, has already been affected by erosion and flooding, which is expected only to worsen in coming decades. The unofficial vote was 89–78 in favour of moving, but there are as yet no funds to pay for relocation.

Hidden Zika risk

A medical report from Brazil confirms that Zika virus can be transmitted through a blood transfusion. Two people received platelet transfusions in January from a blood donor who showed no symptoms at the time of donation, but who was later found to have been infected with Zika virus. The two recipients harboured Zika virus RNA that was genetically related to the virus found in the donor, although they showed no symptoms of infection. The findings add to evidence showing another mode of transmission for Zika, which has been shown to be passed from mother to child *in utero*, and between sexual partners, say the doctors who reported the case on 17 August (I. J. F. Motta *et al. N. Engl. J. Med.* http://doi.org/bppb; 2016).

RESEARCH

Free space for all

NASA announced on 16 August that it is granting free public access to any published research funded by the agency. Research data and peer-reviewed publications by NASA-funded scientists will be available for download and reading on the agency's PubSpace portal within one year of publication. PubSpace was created in response to a 2013 request by the White House Office of Science and Technology Policy. Other US agencies, including the National Institutes of Health and the Food and Drug Administration, are making their research available through the same portal.

Patent thriller

Fresh accusations have rekindled the battle over who invented the potentially lucrative CRISPR–Cas9 gene-editing technology. In an e-mail released by the University of California (UC) on 15 August as part of a pending patent case, Shuailiang Lin, a former student at the Broad Institute in Cambridge, Massachusetts, claims that, contrary to the Broad's claims, the lab built its gene-editing technique on the back of discoveries made at UC Berkeley. The Broad denied Lin's claims and notes that he used the same e-mail to apply for a job at UC Berkeley. The letter was first reported by the *MIT Technology Review*.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

POLICY

Greener lorries

The US government announced stricter fuel-economy standards for lorries, buses and vans on 16 August. Although heavy-duty vehicles make up only about 5% of traffic in the United States, they account for more than 20% of fuel consumption in the transport sector, and contribute a similar proportion of carbon dioxide emissions. The standards for new vehicles will be introduced gradually, becoming tighter each year over the next decade.

EVENTS

Maiden voyage

The world's largest aircraft made its maiden flight in the United Kingdom on 17 August. The 92-metre-long *Airlander 10* (**pictured**) incorporates lighter-than-air technology, combining characteristics of an airship and an aeroplane. It took off from Cardington Airfield in Bedfordshire and performed a circuit of the area before landing safely 20 minutes later. Its developer, Hybrid Air Vehicles in Bedford, says that the hybrid can travel at a cruise speed of 148 kilometres per hour and stay airborne for up to 5 days. The flight marked the start of 200 hours of flight testing. *Airlander 10* is intended to be used for surveillance, communication, aid delivery and even passenger travel.

DARREN STAPLES/REUTERS



Record hot July

July this year was the warmest month since systematic global temperature records began in 1880. It was 0.1 degrees warmer than the previous hottest Julys, in 2015, 2011 and 2009, according to a monthly analysis of global temperatures by scientists at NASA's Goddard Institute for Space Studies in New York City. July was the tenth record-setting warm month in a row, so 2016 looks set to end up as the warmest year in 136 years of modern record-keeping.

Mind the data

Psychologists can now share their data and early results with colleagues before formal publication. On 15 August, a comparative study of personality traits in 8,600 US students was the first paper to be deposited in PsyArXiv, a new preprint server for the psychological sciences (K. S. Corker and B. Donnellan Preprint at PsyArXiv http://osf.io/xeg7y; 2016). Following the example of the successful physics server arXiv, similar online repositories were launched earlier this year for the social sciences and for engineering. A preprint service for chemists will be up and running soon.

PEOPLE

Berkeley head quits

Nicholas Dirks has resigned as chancellor of the University of California, Berkeley. He did not give a specific reason for his decision, announced on 16 August, but he had been criticized for how he managed the university budget and handled allegations of sexual harassment against faculty members and staff — including complaints filed last year against the astronomer Geoffrey Marcy. Marcy stepped down from his position in October 2015, after the complaints came to light.

Russian gambit

On 19 August, Russian President Vladimir Putin appointed church historian Olga Vasilyeva as the country's science and education minister. Vasilyeva succeeds Dmitry Livanov, who will become presidential envoy for trade and economic relations with Ukraine. During his four-year term as minister, Livanov oversaw a radical overhaul of the Russian Academy of Sciences, Russia's main basic-research organization. Vladimir Ivanov, a vice-president of the academy, welcomed Livanov's replacement. Vasilyeva, formerly in charge of public education in religion and history in the presidential administration, told Russian news agency Interfax that religion will not interfere with her work as education and science minister. See http://go.nature.com/2c1zbq1 for more.

Smallpox fighter

Donald Ainslee Henderson, head of the successful campaign to wipe out smallpox, died on 19 August, aged 87. Henderson headed the World Health Organization's global smallpox eradication campaign between 1966 and 1977. The experience caused him to question the feasibility of other disease-eradication efforts. Smallpox remains the only human disease ever to be wiped out, although decades-long efforts to eliminate polio and Guinea-worm disease may now be nearing completion.

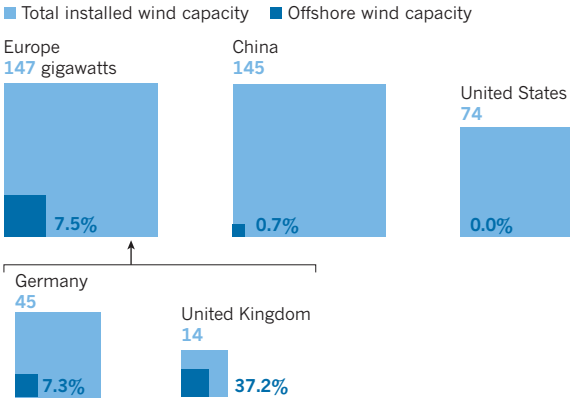
NATURE.COM
For daily news updates see:
www.nature.com/news

TREND WATCH

The UK government's backing of a planned 1.8-gigawatt wind farm off the Yorkshire coast cements Britain's position as the world's leading producer of offshore wind power. More than 90% of global capacity is installed off the coasts of 11 European countries. Most of the remaining 10% is installed off China. In 2015, offshore wind accounted for 24% of total wind-power installations in the European Union. Globally, offshore wind represents only about 3% of installed wind capacity.

EUROPE'S PUSH FOR OFFSHORE WIND

Europe leads the world in installed offshore wind power, with the United Kingdom and Germany contributing the most.



SOURCE: GWEC

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

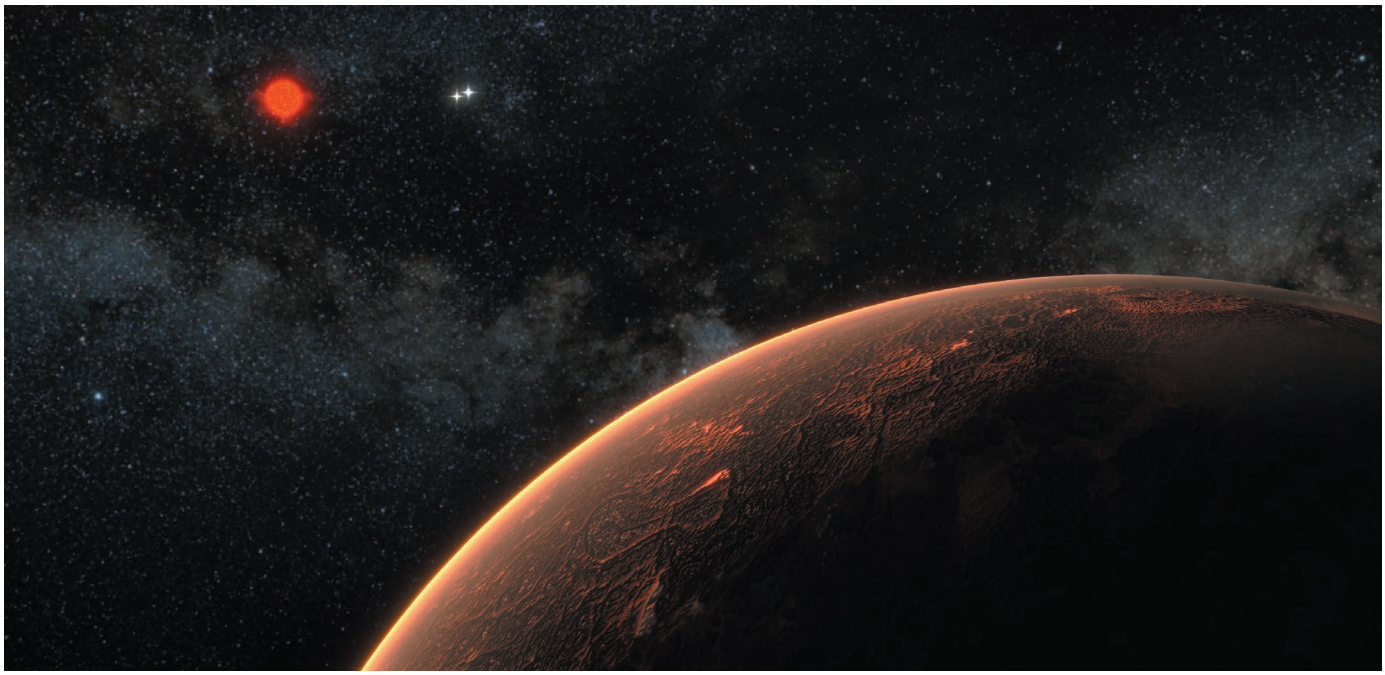
NEWS IN FOCUS

PATENTS Personalized medicine takes hit from US Supreme Court **p.382**

PARTICLE PHYSICS Labs around the world vie to build the next LHC **p.383**

SCIENTIFIC LEGACY The highs and lows of Obama's presidency **p.385**

MEDICINE Non-profit claims secret to cut-price drug discovery **p.388**



RICARDO RAMIREZ

The newly discovered planet orbits Proxima Centauri every 11.2 days.

ASTRONOMY

Nearby star hosts planet

Earth-sized world orbiting Proxima Centauri could harbour water — and life.

BY ALEXANDRA WITZE

Proxima Centauri, the star closest to the Sun, has an Earth-sized planet orbiting it at the right distance for liquid water to exist. The discovery, reported this week in *Nature*[1], fulfils a longstanding dream of science-fiction writers — a potentially habitable world that is close enough for humans to send their first interstellar spacecraft to.

"The search for life starts now," says Guillem Anglada-Escudé, an astronomer at Queen Mary University of London and leader of the team that made the discovery.

Humanity's first chance to explore this nearby world may come from the recently announced Breakthrough Starshot initiative, which plans to build fleets of tiny laser-propelled interstellar probes in the coming decades. Travelling at 20% of the speed of light, they would take about 20 years to cover the 1.3 parsecs from Earth to Proxima Centauri.

Proxima's planet is at least 1.3 times the mass of Earth. The planet orbits its red-dwarf star — much smaller and dimmer than the Sun — every 11.2 days. "If you tried to pick the type of planet you'd most want around the type of star you'd most want, it would be this," says David Kipping, an astronomer at Columbia University in New York City. "It's thrilling."

GRAVITATIONAL HINTS

Earlier studies had hinted at the existence of a planet around Proxima. Starting in 2000, a spectrograph at the European Southern Observatory (ESO) in Chile looked for shifts in starlight caused by the gravitational tug of an orbiting planet. The measurements suggested that something was happening to the star every 11.2 days. But astronomers could not rule out whether the signal was caused by an orbiting planet or another type of activity, such as stellar flares.

In January 2016, Anglada-Escudé and his colleagues launched a campaign to nail down the suspected Proxima planet. ESO granted their request to observe using a second planet-hunting instrument, on a different telescope, for 20 minutes almost every night between 19 January and 31 March. "As soon as we had 10 nights it was obvious," Anglada-Escudé says.

The team dubbed the work the 'pale red dot' campaign, after the famous 'pale blue dot' photograph taken of Earth by the Voyager 1 spacecraft in 1990. Because Proxima is a red-dwarf star, the planet would appear reddish or orangeish, perhaps bathed in light similar to the warm evening tints of Earth.

Although the planet orbits at a distance that would permit liquid water, other factors ▶

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

might render it unlivable. It might be tidally locked — meaning that the same hemisphere always faces the star, which scorches one side of the planet while the other remains cool. The active star might occasionally zap the planet with destructive X-ray flares. And it's unclear whether the planet has a protective, life-friendly atmosphere.

Proxima belongs to the triple-star system Alpha Centauri. In 2012, a *Nature* paper reported that an Earth-mass planet orbited another member of that stellar trio, Alpha Centauri B[2]. That result has now mostly been dismissed[3,4], but exoplanet specialists say that the Proxima claim is more likely to hold up. "People call me Mr Sceptical, and I think this result is more robust," says Artie Hatzes, an astronomer at the Thuringian State Observatory in Tautenburg, Germany.

This time, the combination of new observations and older measurements dating back to 2000 increases confidence in the finding, Anglada-Escudé's team argues. "It's stayed there robustly in phase and amplitude over a very long time," says team member Michael Endl, an astronomer at the University of Texas at Austin. "That's a telltale sign of a planet." The data also contain hints that a second planet might exist, orbiting Proxima somewhere between every 100 and 400 days.

The researchers now hope to learn whether the Proxima planet's pass across the face of its star can be seen from Earth. Such a 'transit' could reveal whether the planet has an atmosphere. A team led by Kipping has been independently looking for transits around Proxima, and is frantically crunching its data in search of a signal.

The discovery of the Proxima planet comes at a time of growing scientific interest in small planets around dwarf stars, says Steinn Sigurdsson, an astrophysicist at Pennsylvania State University in University Park. NASA's Kepler space telescope has shown that rocky planets are common around such stars, which themselves are the most common type of star in the Galaxy. "This is a total vindication of that strategy," he says.

One day, the Proxima planet might be seen as the start of a new stage in planetary research. "It gives us the target and focus to build the next generation of telescopes and one day maybe even get to visit," says Kipping. "It's exactly what we need to take exoplanetary science to the next level." ■ SEE NEWS & VIEWS P.408

1. Anglada-Escudé, G. *et al. Nature* http://dx.doi.org/10.1038/nature19106 (2016).
2. Dumusque, X. *et al. Nature* **491,** 207–211 (2012).
3. Hatzes, A. P. *Astrophys. J.* **770,** 133 (2013).
4. Rajpaul, V., Aigrain, S. & Roberts, S. *Mon. Not. R. Astron. Soc.* **456,** L6–L10 (2016).

INTELLECTUAL PROPERTY

Personalized medicine takes hit

US Supreme Court decisions seem to drive patent rejections.

BY HEIDI LEDFORD

Rejections of US patents in categories related to personalized medicine have spiked after Supreme Court decisions tightened the rules for such claims, an analysis of more than 39,000 patent applications reveals.

The data, presented on 11 August at the Intellectual Property Scholars Conference in Stanford, California, address patent applications in eight categories that commonly include personalized-medicine patents. They show that following a key Supreme Court decision in 2012, the US Patent and Trademark Office (USPTO) was nearly four times more likely to deem subjects of such applications unpatentable — and applicants were less than half as likely to overcome those rejections.

> ***"Personalized medicine is probably the poster child of what we think should be incentivized by patents."***

"The change in office actions was absolutely striking," says Nicholson Price, who studies intellectual property at the University of Michigan Law School in Ann Arbor. "The data are very clear that the patent office has changed its behaviour."

Over the past decade, the Supreme Court has used a series of patent cases to clarify what the USPTO should consider patentable. Natural phenomena and abstract ideas, for example, are not patentable, according to section 101 of the US patent code. The court has attempted to distinguish between these categories and true inventions.

Two of those Supreme Court cases touched directly on the biomedical industry. In 2012, the *Mayo Collaborative Services* v. *Prometheus Laboratories, Inc.* decision struck down two patents on medical diagnostics, and in the 2013 *Association for Molecular Pathology* v. *Myriad Genetics* ruling, the court threw out patents on gene sequences used to assess cancer risk. In the wake of those decisions, many lawyers predicted that patents on inventions that are important to personalized medicine — particularly, diagnostic tests that could match individuals to a particular therapy — would be hard to come by, potentially driving away investors.

Legal scholar Bernard Chao of the University of Denver in Colorado decided to find out just how big the impact has been. Chao sifted through around 85,000 records of USPTO actions taken on more than 39,000 patent applications, and sorted out those that had been rejected for not meeting the requirements of section 101.

He found that, last year, 22.5% of those patent-office actions were rejections because of section 101, compared with only 5.5% in 2011, the year before the *Mayo* decision. Applicants were also less likely to overcome those rejections in the wake of the *Mayo* decision: before *Mayo*, 70.7% of the section 101 rejections were successfully overcome. After *Mayo*, that proportion dropped to 29.7%.

But Chao notes that there are caveats to his analysis: the categories he examined omit some personalized-medicine patents and contain other kinds of patents as well. In the future, he hopes to take a closer look at individual patent applications, and to learn more about whether certain applications are more likely to get through than others.

Those analyses will be key to finding out how patent applicants are adapting to the new requirements, says Price. "Patent attorneys are clever," he says, and may have learnt how to construct their patents to avoid conflict with the recent decisions.

Others have documented a clear effect of the Supreme Court's patent decisions on software patent applications. But some have cheered that change, Chao adds. Software patents are controversial, and some scholars have argued that such patents cause more harm to the industry than help it. Personalized-medicine patents, however, tend to get more support: "Personalized medicine is probably the poster child of what we think should be incentivized by patents."

Ultimately, it will be difficult to unravel what impact the patent decline is having on the personalized-medicine industry, cautions Arti Rai, a legal scholar at Duke University in Durham, North Carolina. The sector is facing challenges from several sides: the US Food and Drug Administration has proposed tougher regulations, and insurance companies have been reluctant to pay for new diagnostic tests.

"Diagnostics start-ups are not in a good space right now, that's clear," Rai says. "But how much of that is due to *Mayo* is less clear." ■

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



HAROLD CUNNINGHAM/GETTY

The Large Hadron Collider has not yet found phenomena outside the standard model of particle physics.

PHYSICS

Who will build the next LHC?

Uncertainty in particle physics complicates collider plans.

BY ELIZABETH GIBNEY

It was a triumph for particle physics — and many were keen for a piece of the action. The discovery of the Higgs boson in 2012 using the world's largest particle accelerator, the Large Hadron Collider (LHC), prompted a pitch from Japanese scientists to host its successor. The machine would build on the LHC's success by measuring the properties of the Higgs boson and other known, or soon-to-be-discovered, particles in exquisite detail.

But the next steps for particle physics now seem less certain, as discussions at the International Conference on High Energy Physics (ICHEP) in Chicago on 8 August suggest. Much hinges on whether the LHC unearths phenomena that fall outside the standard model of particle physics — something that it has not yet done but on which physicists are still counting — and whether China's plans to build an LHC successor move forward.

When Japanese scientists proposed hosting the International Linear Collider (ILC), a group of international scientists had already drafted its design. It would collide electrons and positrons along a 31-kilometre-long track, in contrast to the LHC, which collides protons in a 27-kilometre-circumference circular track at Europe's particle-physics laboratory, CERN, near Geneva, Switzerland.

CLEANER COLLISIONS

Because protons are composite particles made of quarks, collisions create a mess of debris. The particles that would be used in the ILC, by contrast, are fundamental and so provide cleaner collisions more suited to precision measurements, which could reveal deviations from expected behaviour that point to physics beyond the standard model.

For physicists, the opportunity to carry out detailed studies of the Higgs boson and the heaviest, 'top' quark, the second most-recently discovered particle, is reason enough to build the facility. Japan's Ministry of Education, Culture, Sports, Science and Technology (MEXT) was expected to make a call on whether to host the project — which could begin experiments around 2030 — in 2016. But the panel advising MEXT indicated last year that opportunities to study the Higgs boson and the top quark would not on their own justify building the ILC, and that it would wait until the end of the LHC's first maximum-energy run — scheduled for 2018 — before making a decision.

That means the panel is not yet convinced by the argument that the ILC should be built irrespective of what the LHC finds, says Masanori Yamauchi, director-general of Japan's High Energy Accelerator Research Organization (KEK) in Tsukuba, who sat on an ICHEP panel at a session on future facilities. "That's the statement hidden under their statement."

If the LHC discovers new phenomena, these would be further fodder for ILC research — and would strengthen the case for building the high-precision machine.

US physicists have long backed the construction of a linear collider. And Yamauchi says that a joint MEXT and US Department of Energy group is discussing ways to reduce the ILC's costs, which are now estimated at US$10 billion. A reduction of around 15% is feasible — but Japan will need funding commitments from other countries before it formally agrees to host, he added.

CHINESE COMPETITOR

Snapping at Japan's heels is a Chinese team. In the months after the Higgs discovery, physicists led by Wang Yifang, director of the Institute of High Energy Physics in Beijing, floated a plan to host a collider in the 2030s, also partially funded by the international community and focused on precision measurements of the Higgs and other particles. Circular rather than linear, this 50–100-kilometre-long electron–positron smasher would not reach the energies of the ILC. But it would require the creation of a tunnel that could allow a proton–proton collider — similar to the LHC, but much bigger — to be built at a hugely reduced cost.

Wang and his team this year secured around 35 million yuan (US$5 million) in funding from China's Ministry of Science ▶

MORE ONLINE

FIELD MUSEUM LIBRARY/GETTY

TOP NEWS



CRISPR provides 'evo-devo' test bed for weird creatures go.nature.com/2bsbebe

MORE NEWS

- Brain's chemical signals seen in real time go.nature.com/2byhttd
- Most radically re-wired bacterial genome created go.nature.com/2bdirol
- Church historian is Russia's new science minister go.nature.com/2c1zbq1

NATURE PODCAST

Earth-like planet close by; dietary restriction combats ageing syndrome; drugs for neglected diseases nature.com/nature/podcast

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

and Technology to continue research and development for the project, Wang told the ICHEP session. Last month, China's National Development and Reform Commission turned down a further request from the team for 800 million yuan, but other funding routes remain open, Wang said. The team now plans to focus on raising international interest in the project.

By affirming worldwide interest in Higgs physics, the Chinese proposal bolsters the case for building the ILC, says Yamauchi. But if it goes ahead, it could eventually drain international funding from the ILC. "It may have a negative impact," he says.

SUPER-LHC

In the future, the option to use China's electron–positron collider as the basis for a giant proton collider could interfere with CERN's own plans for a 100-kilometre-circumference circular machine that would smash protons together at more than 7 times the energy of the LHC (see 'World of colliders'). Until the mid-2030s, CERN will be busy with an upgrade that will raise the intensity — but not the energy — of the LHC's proton beam. By that time, China might have a suitable tunnel that could make it harder to get backing for this 'super-LHC'.

At ICHEP, Fabiola Gianotti, CERN's director-general, floated an interim idea: souping up the energy of the LHC beyond its current design by installing a new generation of superconducting magnets by around 2035. This would provide a relatively modest boost in energy — from 14 teraelectronvolts (TeV) to 20 TeV — that would have a strong science case if the LHC finds new physics at 14 TeV, said Gianotti. Its $5-billion price tag could be paid for out of CERN's normal budget.

For decades, successive facilities have found particles predicted by the standard model, and neither the LHC nor any of its proposed successors is guaranteed to find new physics.

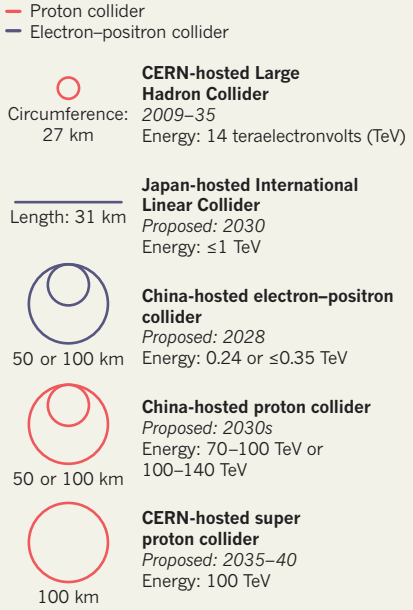
Questions from attendees at the ICHEP session revealed some soul-searching, including a plea to reassure young high-energy physicists about the field's future, and contemplation of whether money would be better spent on other approaches rather than ever-bigger accelerators.

The United States is betting on neutrinos, fundamental particles that could reveal physics beyond the standard model, not colliders. The Fermi National Accelerator Laboratory (Fermilab) in Batavia, Illinois, hopes to become the world capital of neutrino physics by hosting the $1-billion Long-Baseline Neutrino Facility, which will beam neutrinos to a range of detectors starting in 2026.

Funding will require approval from US Congress in 2017. But at the ICHEP session, Fermilab director Nigel Lockyer was confident: "We are beyond the point of no return. It is happening." ■

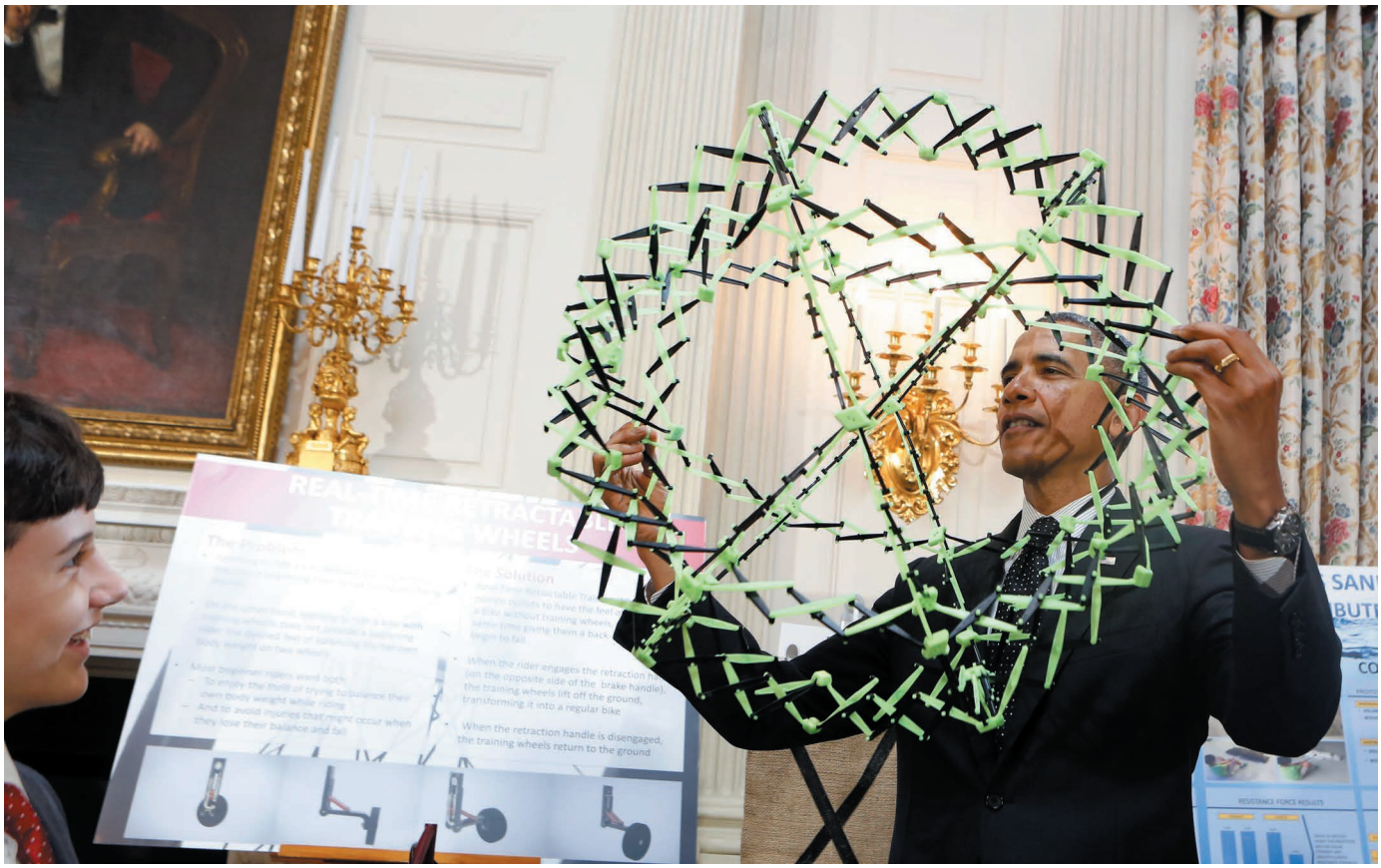
WORLD OF COLLIDERS

Physicists around the world are designing a range of particle colliders that are much bigger than the Large Hadron Collider at CERN, Europe's particle-physics laboratory.



© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

AUDE GUERRUCCI/POOL/CORBIS/GETTY



Barack Obama embraced science, but his policies get mixed reviews.

POLITICS

Grading America's 'scientist-in-chief'

US President Barack Obama sought to map the brain, cool the climate and chart a path to Mars. As he prepares to leave office, Nature *looks back at the scientific highs and lows of his presidency.*

BY HEIDI LEDFORD, RICHARD MONASTERSKY, JEFF TOLLEFSON & ALEXANDRA WITZE

BETTING BIG ON BIOMEDICAL SCIENCE

When president-elect Barack Obama chose physicist John Holdren as his top science adviser in December 2008, some biomedical researchers worried that the pick signalled a White House bias towards physical science.

Obama quickly put those fears to rest. Within weeks of his inauguration, he had overturned restrictions on research using embryonic stem cells. He has gone on to launch major initiatives to map the brain, develop personalized medical treatments and cure cancer.

But faced with a penny-pinching Congress, Obama's strong support for biomedical science has not translated into significant funding gains for the US National Institutes of Health (NIH). The agency has seen the purchasing power of flat research budgets eroded by inflation (see 'Budget battles'). "The life sciences were a significant priority for the Obama administration," says Gregory Petsko, a biochemist at Weill Cornell Medical College in New York City. "But with Congress being the way that it is, there was a limit to what Obama could do as far as increasing support of biomedical research."

It is the big initiatives that will probably form Obama's lasting biomedical legacy, says Benjamin Corb, head of public affairs at the American Society for Biochemistry and Molecular Biology in Rockville, Maryland. In 2013, Obama announced the Brain Research Through Advancing Innovative Neurotechnologies (BRAIN) initiative to map the human brain. In 2015, he unveiled the Precision Medicine Initiative, which includes an ambitious study of health records and genomic information from one million people in the United States. And in January, he introduced the Cancer Moonshot, a US$1-billion proposal to double the pace of cancer research in five years.

NIH director Francis Collins, who led the Human Genome Project in the 1990s, likens Obama to a player who scores three goals in the same game: "I said to him, basically, 'Mr President, you have achieved a hat-trick.'"

But such programmes may come at a cost to basic research funding, even as they draw attention to areas of science that may be overlooked or underfunded. "These big initiatives tend to cast a really large shadow," says Corb. "They can overshadow some of those basic research needs."

And it's not clear whether Obama's major initiatives will survive under the next president. Democratic presidential candidate ▶

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Hillary Clinton has said that she would continue the Cancer Moonshot initiative. She also supports Alzheimer's disease research, which bodes well for the BRAIN initiative if she is elected, Corb says. Republican candidate Donald Trump has no clear policy on biomedical research.

But the next president won't be making that decision alone. Patient advocates drive major changes in biomedical research priorities and funding over time, and will probably ensure that Obama's big-science initiatives continue, says Mary Woolley, president of the science-advocacy organization Research!America in Arlington, Virginia. "Determined advocates are not going to take 'no' for an answer," she says. "They'll be the ones that bridge administrations."

SPACE RACE STALLS

Obama tried to shake up the US space programme, including NASA's long-standing plan to send people to Mars. But nearly eight years — and a series of U-turns — later, he has little to show for his effort.

"Where NASA is today is really not all that different from where it was during the last presidential transition," says Marcia Smith, a space-policy analyst in Arlington, Virginia, who runs SpacePolicyOnline.com.

A crewed Mars mission remains two decades away. Its schedule is constrained by the funding available to develop the necessary hardware — a new heavy-lift rocket and crew capsule to sustain astronauts in deep space. That is almost exactly the situation NASA was in eight years ago, bar one detail: Obama ditched the Moon as a first stop for astronauts on their way to Mars.

That decision, in February 2010, stunned NASA, Congress and space-policy experts. Obama cancelled the Constellation programme, which his predecessor George W. Bush created to send US astronauts back to the Moon in preparation for an eventual Mars trip. Two months later Obama announced a different course: astronauts would visit a yet-to-be-chosen asteroid before heading off to the red planet. The White House did not consult Congress on the switch, angering powerful members who represent space-industry employees in states such as Florida, Texas and California. "The hostility created by the way the Obama administration rolled that out still lingers in Congress," says Smith.

The decision also alienated traditional US space partners such as Europe and Japan, says Scott Pace, director of the Space Policy Institute at George Washington University in Washington DC. "Little to no weight was given to the international implications of the decision to abandon efforts to lead an international return of humans to the lunar surface," he says.

NASA was forced to modify its Mars plan in 2013, when it became clear that it did not have the technology to support astronauts in deep space. The White House introduced a controversial stopgap measure: instead of a crewed mission to visit an asteroid, a robot would drag an asteroid near the Moon where astronauts could then visit it.

Asteroid scientists have roundly denounced the plan, but it is moving forward despite slipping schedules and ballooning costs. The hardware for crewed deep-space journeys is also at risk of schedule and budget delays, the Government Accountability Office said last month. The heavy-lift rocket is scheduled for its first test flight in November 2018, while the crew capsule's is set for August 2021.

Obama extended US participation in the International Space Station for four more years, to 2024 — a move generally acclaimed by scientists. And he oversaw the shutdown of the space-shuttle programme, a process begun by Bush. After the last shuttle, *Atlantis*, flew in July 2011, the United States turned to Russia to buy rides to orbit for its astronauts.

NASA is relying on commercial companies to fly equipment and — eventually — astronauts to the space station. The first commercial cargo flights began in 2012, and the first astronauts are scheduled to fly aboard commercial spaceships no earlier than 2017.

Many critics see NASA's human-spaceflight programme as adrift. Eileen Collins, a former space-shuttle commander, told the Republican National Convention in July that the agency needs "visionary leadership again".

Scientists grumble about the relative lack of flagship missions in development. One of the biggest, a proposed mission to Jupiter's moon Europa, has been pushed through not by the White House or NASA, but by a Republican congressman from Texas who is enamoured with the idea of life on icy worlds.

UNEVEN PROGRESS ON SCIENTIFIC INTEGRITY

Many researchers who watched Obama's inauguration in 2009 were thrilled by his pledge to "restore science to its rightful place". But scientists and legal scholars say that, in many ways, Obama has failed to live up to that lofty promise.

In general, government researchers have enjoyed more freedom — and endured less political meddling — than they did under the previous president, George W. Bush. Bush's administration was accused of muzzling or ignoring scientists on subjects ranging from stem cells to climate change.

In March 2009, Obama instructed agencies to develop policies to reduce political interference and increase transparency about the research used in policy decisions. And when the Union of Concerned Scientists (UCS) surveyed federal researchers in 2015, most said that their agency adhered to its scientific-integrity policy.

But critics say that Obama's White House has not shied away from exerting political influence over science.

In 2011, the Environmental Protection Agency (EPA) sent a proposal to the White House that would strengthen controls on ozone pollution, based on guidance from its scientific advisers. But Obama directed the agency to withdraw the plan, citing the cost of the stricter limits at a time when the economy was still recovering from a recession.

And that same year, Health and Human



LUCY NICHOLSON/REUTERS

Obama shut down the space shuttle, and encouraged the growth of a commercial space industry.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Services secretary Kathleen Sebelius overruled the Food and Drug Administration's finding that the emergency contraceptive 'Plan B One-Step' was safe to dispense over the counter for all women and girls.

In both cases, science eventually won out: the EPA approved stronger ozone standards in 2015, and the FDA approved unrestricted sales of Plan B in 2013 after judges ruled against the agency. Nevertheless, these examples show how political considerations have sometimes trumped scientific ones during Obama's tenure, says Lisa Heinzerling, a law professor at Georgetown University in Washington DC.

"There are structures in place that threaten scientific integrity and encourage the injection of politics into matters that are supposed to be scientific or technical," says Heinzerling, who worked at the EPA for two years under Obama.

Science advocates are concerned about how political influence shapes science behind closed doors at the White House. The president's Office of Management and Budget, which reviews proposals for new rules and regulations, can make substantial changes or kill a policy without explaining why. "In some cases, the White House is messing around, and it's not doing it transparently," says Wendy Wagner, a law professor at the University of Texas at Austin.

The recent UCS survey revealed room for improvement at several agencies. Nearly half the scientists at the Centers for Disease Control and Prevention said that their agency gave too much weight to political interests; that proportion rose to 73% at the Fish and Wildlife Service. And less than 60% of scientists at the four agencies surveyed said they could openly express concerns about the work of their employer without fear of retaliation.

"We have a lot of new policies and procedures in place that are tremendously beneficial," says Gretchen Goldman, a UCS analyst who led the study. "But what we're finding is that there's more work to be done."

CLIMATE POLICY HOTS UP

Global warming was one of Obama's top priorities — and one of the most difficult to address, given strong opposition from Republicans in Congress. Yet he managed to help broker a global climate accord and push through regulations to curb greenhouse-gas emissions from cars, trucks and power plants.

"Obama has established a terrific climate legacy," says David Doniger, who directs the climate and clean-air programme at the Natural Resources Defense Council, an advocacy group in New York.

The president's earliest actions capitalized on the global financial crisis. In February 2009, Obama signed economic-stimulus legislation that included nearly $37 billion for clean-energy research and development (R&D) at the Department of Energy. Four months later, with failing car companies seeking a federal bailout, the Obama administration proposed higher fuel-efficiency requirements and the first greenhouse-gas standards for passenger vehicles. The regulations, which took effect in 2012, will nearly double the average fuel efficiency of vehicles by 2025, to around 23 kilometres per litre.

And after his campaign for a comprehensive climate bill failed in 2010, an emboldened Obama used existing laws to issue regulations that curbed greenhouse-gas emissions, bolstered energy-efficiency standards and expanded energy R&D programmes.

"Paris is a major achievement for the world."

But the president's big push on climate came in advance of the United Nations climate summit in Paris in 2015. He committed the United States to reduce emissions by at least 26% below 2005 levels by 2025, and negotiated directly with countries such as China to build support for a global climate agreement. The final version, adopted on 12 December, aims to hold average global temperatures to 1.5–2 °C above pre-industrial levels.

"Paris is a major achievement for the world," says Robert Socolow, a climate scientist at Princeton University in New Jersey. "I don't think it would have happened without Obama."

Yet Obama's domestic achievements could be undone by legal challenges. In February, the US Supreme Court temporarily blocked a federal regulation to reduce emissions from existing power plants. The fate of that rule — the cornerstone of Obama's plan to reduce emissions — could depend on the election in November. The Supreme Court is down one member and the next president will choose a replacement, who could decide whether the climate rule stands.

Some environmental experts say that Obama should have pushed harder for a comprehensive climate bill, rather than settling for piecemeal regulations. "All of these things are actually small bites at the apple that won't achieve meaningful emissions reductions over time," says Catrina Rorke, director of energy policy at the R Street Institute, a conservative think tank in Washington DC.

Others criticize Obama for encouraging a vast expansion of domestic oil and gas development, even as he sought to wean the country off coal and curb its greenhouse-gas emissions. "The administration is still trying to have it both ways," says Stephen Kretzmann, executive director of Oil Change International, an advocacy group in Washington DC.

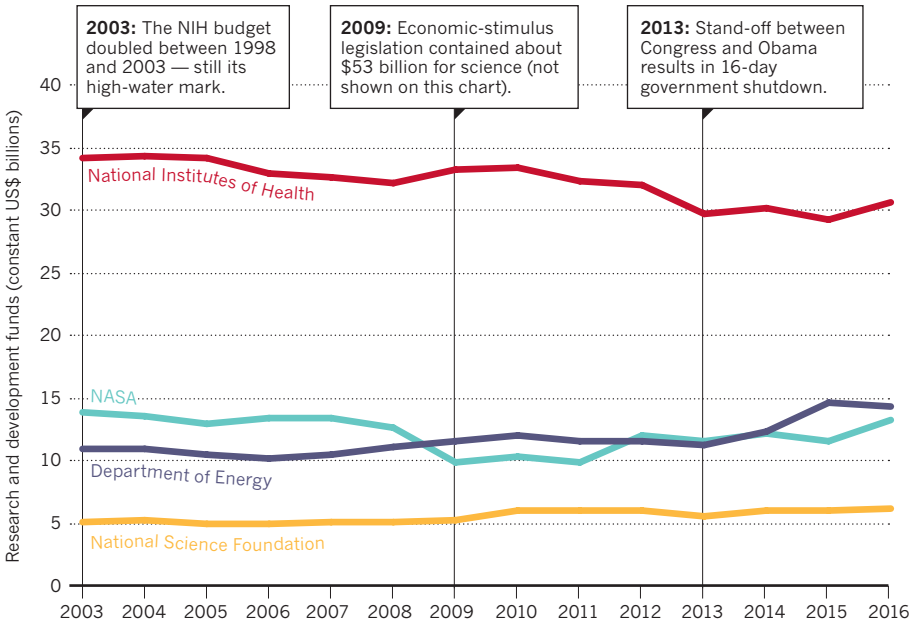
Obama rejected the Keystone XL pipeline, which would have carried oil from the Canadian tar sands to US refineries, and has said that some fossil fuels should be kept "in the ground". But his administration continues to push an 'all-of-the-above' energy strategy that leads to higher production of domestic fossil fuels, Kretzmann says.

Nonetheless, Obama has helped to change the conversation about global warming at home and abroad, says Doniger. "The next president needs to do more," he says, "but did the Obama administration move the ball forward? They sure did." ■

Additional reporting by Sara Reardon

BUDGET BATTLES

US President Barack Obama, who took office in January 2009, pushed to increase funding for science agencies. But Congress often rebuffed his proposals.



SOURCE: AAAS R&D BUDGET AND POLICY PROGRAM

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

BIG PHARMA'S – COST-CUTTING CHALLENGER –

A non-profit organization is proving that drug development doesn't have to cost a billion dollars. Can its model work more broadly?

BY AMY MAXMEN

First, there was the pitching and rolling in an old Jeep for eight hours. Next came the river crossing in a slender canoe. When Nathalie Strub Wourgaft finally reached her destination, a clinic in the heart of the Democratic Republic of the Congo, she was exhausted. But the real work, she discovered, had just begun.

It was July 2010 and the clinic was soon to launch trials of a treatment for sleeping sickness, a deadly tropical disease. Yet it was woefully unprepared. Refrigerators, computers, generators and fuel would all have to be shipped in. Local health workers would have to be trained to collect data using unfamiliar instruments. And contingency plans would be needed in case armed conflict scattered study participants — a very real possibility in this war-weary region.

This was a far cry from Wourgaft's former life as a top executive in the pharmaceutical industry, where the hospitals that she commissioned for trials were pristine, well-resourced and easy to reach. But Wourgaft, now medical director for the innovative Drugs for Neglected Diseases initiative (DNDi), was confident that the clinic could handle the work. She was right. With data from this site and others, the DNDi will next year seek approval for a sleeping-sickness tablet, fexinidazole. It would be a massive improvement on existing treatment options: an arduous regimen of intravenous injections, or a 65-year-old arsenic-based drug that can be deadly.

The DNDi is an unlikely success story in the expensive, challenging field of drug development. In just over a decade, the group has earned approval for six treatments, tackling sleeping sickness, malaria, Chagas' disease and a form of leishmaniasis called kala-azar. And it has put another 26 drugs into development. It has done this with US$290 million — about one-quarter of what a typical pharmaceutical company would spend to develop just one drug. The model for its success is the product development partnership (PDP), a style of non-profit organization that became popular in the early 2000s. PDPs keep costs down through collaboration — with universities, governments and the pharmaceutical industry. And because the diseases they target typically affect the world's poorest people, and so are neglected by for-profit companies, the DNDi and groups like it face little competitive pressure. They also have lower hurdles to prove that their drugs vastly improve lives.

Now, policymakers are beginning to wonder whether their methods might work more broadly. "For a long time, people thought about R&D as so complicated that it could only be done by the biggest for-profit firms in the world," says Suerie Moon, a global-health researcher at the Harvard T.H. Chan School of Public Health in Cambridge, Massachusetts, who studied PDPs and joined the DNDi's board of directors in 2011. "I think we are at a point today where we can begin to take lessons from their experience and begin to apply to them non-neglected disease," she says.

In that vein, the DNDi has started research on alternatives to pricey drugs for hepatitis C, and is spearheading an effort to create antibiotics for drug-resistant infections, a problem that pharmaceutical companies have been slow to contend with. If successful, the work could challenge standard assumptions about drug development, and potentially rein in the runaway price of medications. "We can't match our financial figures one to one," says executive director Bernard Pécoul. "But we believe that DNDi can demonstrate that a different model is possible for R&D."

THE PIPELINE

When medical charity Médecins Sans Frontières (MSF; also known as Doctors without Borders) won the Nobel Peace Prize in 1999, its members decried the lack of lifesaving drugs for diseases of the poor, and used the Nobel prize money to kick-start the DNDi. Pécoul, a soft-spoken Frenchman who had been with MSF for 20 years, took the helm when the initiative launched in Geneva, Switzerland, in 2003. Pharmaceutical executives were sceptical. Drug development is an expensive, complex, decade-long endeavour. "In the early days, we saw DNDi as a bit amateurish," recalls François Bompart, a medical director at the Paris-based drug company Sanofi. "We thought, they cannot be serious."

Pécoul and his team started with a safe project. In 2001, the World Health Organization had called for malaria drugs that combined ingredients to slow the spread of resistance to the single best available agent, artemisinin. But the poverty of most people who need malaria drugs meant that the private sector had little incentive to create and test such combination therapies. Pécoul contacted Sanofi, which owned two malaria treatments: one based on artemisinin, and the other on the

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



DNDi medical director Nathalie Strub Wourgaft examines a child in Sudan.

NEIL BRANDVOLD/DNDI

slower-acting amodiaquine. He proposed a deal in which the DNDi would pay for and run clinical trials on a pill that combined the two drugs. In return, Sanofi would not patent the pill and would sell an adult course of treatment for no more than $1, half that for children. "To me it sounded very aggressive and not reasonable, since the two drugs separately were two to three times that," says Bompart.

But Pécoul convinced Sanofi that the move would be good for the company's public image. He also compromised, allowing Sanofi to stipulate that it could reach the low price gradually. As it turned out, by the time the pills were approved in 2007, manufacturing costs had come down far enough for the company to meet the target price right out of the gate. Hundreds of millions of pills have since been distributed in Africa. All told, the project cost the DNDi about $14 million, a tiny sum in the world of drug development. It has since replicated the process to develop other combination therapies (see 'Discount drugs').

Although they improve on existing therapies, some of these combinations remain inadequate. The DNDi's sleeping-sickness therapy NECT, for example, reduces a standard treatment from 56 intravenous infusions to 14. That is still problematic in affected countries: clean needles can be hard to come by, and long hospital stays are often impossible. People need a pill.

Drug development from scratch is arduous and expensive. It begins with experiments on hundreds of thousands of chemicals in the lab, looking for one that kills a pathogen without harming the host. The DNDi does not have a laboratory, so it does this through collaborations. It searches for promising leads in compound libraries generated by biotechnology and pharmaceutical companies. Many firms are willing to share access to these precious libraries because the diseases that the DNDi targets will not result in blockbuster drugs, so it is not infringing on their turf. The DNDi then contracts high-throughput screening centres, such as those at the Institut Pasteur Korea in Seongnam and the University of Dundee, UK, to test them out. "We use the same technique that pharma does," says Rob Don, director of discovery and preclinical research at the DNDi, "but we do it for less."

In 2007, such efforts identified fexinidazole, a compound that had shown promise against single-celled parasites but was pulled from development before reaching clinical trials. The DNDi turned it into a tablet, and passed it to its clinical-development team two years later.

The DNDi approached Sanofi again and promised to take care of trials if the company could file for regulatory approval. Sanofi warned that human trials would not be easy, because sleeping sickness is not common and people who get it tend to live in remote, unstable regions. But with the existing therapies being so dreadful, Wourgaft argued that any improvements from fexinidazole would be clear. "The delta between what we bring and what exists is huge. You don't need a magnifying glass on thousands of patients to see it." She set up multiple small trial sites in the Democratic Republic of the Congo and the Central African Republic and pooled their data.

> "WE USE THE SAME TECHNIQUE THAT PHARMA DOES, BUT WE DO IT FOR LESS."

CLINICAL CHALLENGE

Wourgaft says that the studies were the hardest she has ever run. In addition to logistical challenges, civil war erupted in the Central African Republic shortly after the study launched, and rebel groups repeatedly robbed a clinic there and threatened the Congolese surgeon leading the trial. "I squeeze all my energy into each project," Wourgaft says. "It's as if I'm using forceps to deliver a baby — and the baby is an elephant."

The final trials on fexinidazole conclude this year, and Wourgaft is hopeful that the data will earn regulators' stamp of approval. The project has so far cost the DNDi about $45 million — and it stands to help 21 million people at risk of the disease in Africa. In a few months, Wourgaft will launch another trial, on a completely new oral

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

DISCOUNT DRUGS

The Drugs for Neglected Diseases initiative (DNDi) has produced several drugs in the past decade for a fraction of what pharmaceutical companies are said to spend. Factoring in the cost of failed candidates (not included below), the DNDi estimates that it can develop combination therapies for between US$10 million and $45 million, and make a completely new drug from scratch for $110 million to $170 million.

Drug \| Disease	Research	Early safety and proof-of-concept trials	Late safety and efficacy trials	Access and additional studies	Total cost
COMBINATION THERAPIES					
SSG&PM \| Kala-azar (visceral leishmaniasis)	Not needed for combinations of approved compounds		$10.5m	$2.5m	**$13m**
NECT \| Sleeping sickness			$4m	$3.6m	**$7.6m**
NOVEL DRUGS					
Fexinidazole \| Sleeping sickness	$8.1m	$5m	$32m	$17m*	**$62m**
SCYX-7158 \| Sleeping sickness	$25m	$4.1m	$20.8m	$17m*	**$67m**

*Projected estimates until 2020

SOURCE: DNDI

drug — SCYX-7158 — that may cure people with sleeping sickness in a few days. The DNDi estimates that its development up to approval will cost around $50 million.

BREAKING BILLIONS

For more than three decades, economists at the Tufts Center for the Study of Drug Development in Boston, Massachusetts, have collected proprietary data from pharmaceutical companies, and used it to calculate the average cost of developing a new drug. The most recent estimate is $1.4 billion. This is used to justify exorbitant drug prices — companies must recoup their investments. But many don't think it has to cost that much. Even the chief executive of London-based pharmaceutical giant GlaxoSmithKline, Andrew Witty, has called billion-dollar estimates "one of the great myths of the industry". He attributed the huge sums to spending too much time on failures. Drug candidates can be killed as a result of safety concerns, poor efficacy or profitability worries, and he argued that companies could save money by dropping bad leads sooner. Others say that the figure is inflated by large and excessive trials done to prove that a new drug works just slightly better than an existing one.

By averaging the cost of projects in its portfolio, the DNDi says that it can develop a new drug for between $110 million and $170 million. Like the Tufts estimate, these prices include a theoretical cost of failed projects.

The DNDi admits to enjoying perks that pharma does not have. It keeps overhead costs low because its organization is virtual. The research organizations that it contracts probably charge the group less than they would a for-profit company. The DNDi also relies on scientific consultants who work for low pay because they relish the chance to make lifesaving drugs without considering competitors, investors and marketing. "DNDi gets a lot for free," says Richard Bergström, director-general of the European Federation of Pharmaceutical Industries and Associations in Brussels. "My companies do a lot of pro bono work, and so do universities."

Still, the organization reckons that such in-kind contributions account for just 10–20% of its expenditure. It saves much more through efficient collaboration (avoiding duplicated effort by screening pooled libraries, for example) and a focus on desperately needed drugs. Clinical trials can be smaller, faster and cheaper when the people who run them don't have to struggle to show barely perceptible improvements. And the DNDi kills candidate compounds only if they fail on safety or efficacy — it doesn't have to worry about marketability. By contrast, a few for-profit companies froze candidate drugs for hepatitis C after Gilead Sciences of Foster City, California, brought powerful drugs to the market. "A lot of R&D failures in pharma are commercial rather than scientific," says Don. "We keep going until it gets to market or scientifically fails."

The DNDi has earned respect from the industry, even though its founding organization has been antagonistic to big pharma. "Although DNDi came out of MSF, they don't let ideological viewpoints get in the way of making progress," says Jon Pender, vice-president of government affairs at GlaxoSmithKline. He and others praise Pécoul's skills at negotiation, and the DNDi's pragmatic approach to development challenges.

Policymakers have taken notice, too. Last year, the World Health Organization asked the DNDi to consider antibiotics for drug-resistant infections in the developing world; in May, the initiative announced that it would start the GARD (Global Antibiotic Research and Development) partnership with $2.2 million in seed funding. GARD will start by repurposing and combining existing antibiotics to treat a few diseases, including gonorrhoea and infections in newborn babies. Marja Esveld, a research adviser at the Netherlands ministry of health, is watching it closely. "We are worried about the rising costs of pharmaceuticals," she says, "and so for us, GARD is also a kind of experiment to see if the DNDi model can work for the development of drugs in the Western world."

Not everyone is convinced. Economist Ramanan Laxminarayan, director of the Center for Disease Dynamics, Economics and Policy in Washington DC, says that pharmaceutical companies have an incentive to make antibiotics for multidrug-resistant infections because patients in the United States and Europe will pay to get them — and non-profit organizations cannot hope to compete. Once the drugs exist, he says, subsidies could ensure that they are affordable.

Pécoul disagrees: he doesn't think that subsidies, donations or tiered pricing can ensure accessibility. "We need appropriate products and a sustainable market for those products," he says. That environment has not materialized for other conditions: Gilead's hepatitis C drugs, for example, are listed at more than $74,000 for a course. And their potency against some strains of the virus is questionable, says Pécoul. When he and his team learned about other hepatitis drug candidates being frozen, he launched a project to turn them into treatments that more people could use and afford. They're also attempting to combine existing drugs.

If the group succeeds with this and with antibiotics, it will have shown that its model can be applied to diseases that affect developed countries. "I hope we provide lessons that can be used by others," says Pécoul. But companies won't simply adopt the DNDi's methods, because they do not generate profit. The investors who keep firms alive are concerned with the bottom line. Pécoul says that a transformation would require government involvement and a reorganization of the development process. It would need a system to prioritize what treatments are needed and which companies and organizations could collaborate; and it would require forethought about how the final products would reach those in need. It means shifting away from profit-based incentives to things such as prizes and government funding. Today's profit-driven approach is not only expensive, Pécoul says, it fails huge swathes of the population.

When Wourgaft reflects on the differences between her career in pharma and her work at the DNDi, she thinks not about the cost of research and development, but about the value of a human life. She recalls one trip to a Congolese sleeping-sickness trial site. She sat on a cot beside a woman in the middle of a psychotic episode, and spoke to her desperate husband. Later, she learned that the woman survived because of the DNDi's treatment. "When you see that, you know the value of what you're doing," she tells me. "We are trying to fix diseases that are lethal — this is really serious medicine." ■

Amy Maxmen *is a science journalist in Berkeley, California.*

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

COMMENT

BIOETHICS *Flowers for Algernon* is still powerful 50 years on **p.394**



HISTORY The Victorian female surgeon who had to pose as a man **p.395**

BRAIN DRAIN Venezuela is haemorrhaging its leading researchers **p.396**

OBITUARY Alfred G. Knudson, cancer-genetics pioneer, remembered **p.397**

LEE POWELL/*THE WASHINGTON POST*/GETTY



New York City's High Line park, a transformed former rail line.

Expand the frontiers of urban sustainability

Social equity and global impacts are missing from measures of cities' environmental friendliness, write **David Wachsmuth**, **Daniel Aldana Cohen** and **Hillary Angelo**.

Manhattan skyscrapers, rather than rustic rural towns, are quickly becoming the picture of sustainable living in the twenty-first century. San Francisco, Copenhagen and Singapore each top their regions in the Green City Index (see go.nature.com/2bxjac9). As sites of innovation and economic dynamism, these places exemplify a blend of density and livability that large, prosperous cities in the 'global south', such as Mumbai in India and São Paulo in Brazil, increasingly emulate.

A few decades ago, cities were seen as sustainability problems rather than solutions. Then, as concerns about suburban sprawl, shanty towns and climate change grew, so too did awareness that clustering people in energy-efficient buildings and walkable, shady neighbourhoods makes cities more pleasant to live in and better for the global environment.

But the prevailing model of urban sustainability is too narrow. Although the social, economic and ecological issues behind sustainability problems are regional or global in scale, urban policy usually addresses single ecological ▶

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

issues in individual neighbourhoods. Focusing on dense cities and their affluent areas ignores social movements and their advocacy for quality-of-life issues such as housing and commuting, which have direct ecological consequences. Targeting specific districts ignores the often negative regional and global impacts of local environmental, or 'greening', improvements.

"Even information in 'the cloud' has an environmental impact."

Spatially, sustainability research and policymaking should shift focus from city centres to urban regions and global networks of production, consumption and distribution. Socially, policymakers should incorporate equity into every stage of the urban-policy process, from research to formulation to implementation.

NEIGHBOURHOOD WATCH

From the revitalization of city parks to urban bicycle-sharing programmes, urban sustainability interventions tend to be conceived, implemented and evaluated one municipality or neighbourhood at a time. Yet urban environmental processes occur on much larger scales. Projects that benefit one district may have negative impacts next door.

One example is environmental gentrification. As districts become greener, they become more desirable and expensive. The premiums placed on neighbourhood amenities — such as walkability, public transport and the proximity of parks, farmers' markets and 'greenways' such as hiking trails and bike paths — by residents who can afford to pursue them raise the cost of living.

Social displacement can result. Policies that encourage these improvements tend not to be linked to a broader social-equity agenda, so low- and middle-income residents are forced into peripheral neighbourhoods where population densities are lower, commutes are longer and environmental problems are more common. Many sustainability gains are simply a regressive redistribution of amenities across places.

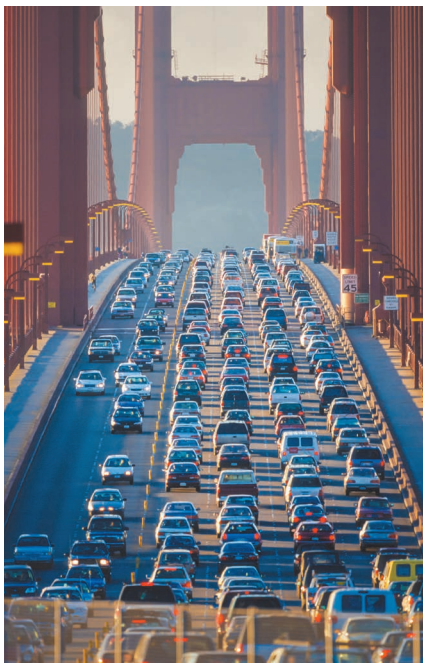
For example, in North American cities such as New York and San Francisco, poor districts have long suffered from the dumping of industrial-waste, low air quality and a lack of green spaces. In recent years, often in response to community activism, policymakers have tried to create shadier streets and more recreational space, to improve public transport and greenway access, and to build mixed-use eco-friendly housing in such neighbourhoods. New York City has made efforts to green East Harlem, western Queens and Red Hook in Brooklyn. Yet poor people are frequently priced out and must move[1].

In Europe, the German city of Freiburg has been internationally recognized for its achievements in renewable energy, public transport, participatory planning and pedestrianized, energy-efficient districts. As the metropolitan region has become more desirable and expensive, more of its workforce has turned to the cheaper suburbs for housing. The city has grown more socially homogenous, while beyond its boundaries commuting has skyrocketed, as have the associated carbon emissions[2]. Greening has come at the expense of community stability and racial and economic diversity, and has undermined regional environmental goals.

These patterns hold around the world. Studies have shown that in several cities, the social costs of climate adaptation fall mainly on disadvantaged groups. Examples include Medellín, Colombia; Jakarta, Indonesia; Dhaka, Bangladesh; and Boston, Massachusetts. Climate-adaptation plans fail to engage poor communities and often recommend relocating them to unsafe areas where they would be more vulnerable to droughts, heat, flooding and disease. Meanwhile, wealthy residents who set the planning agenda benefit from new land-use regulations and protective infrastructure. From Boston to Dhaka, resources earmarked for climate-adaptation are concentrated in wealthy districts and the risks are exacerbated elsewhere[3].

FARTHER AFIELD

Post-industrial cities highlight their sustainability triumphs in terms of building density, extensive public-transport networks and the presence of knowledge-intensive, high-tech firms, all of which drive down locally produced pollution and carbon emissions. But even high-tech workplaces depend on polluting activities elsewhere. Computers and smartphones produce growing global flows of electronic waste that concentrate their toxic by-products — such as the trace amounts of beryllium and mercury in mobile phones — in poor communities in the developing world. Guiyu in China used to be a small rice-growing village, but was transformed in the 1990s into the world's largest processing zone for electronic waste. Local water rapidly became undrinkable[4].



The Golden Gate Bridge in San Francisco.

DAVID MADISON/GETTY

Even information in 'the cloud' has an environmental impact. Data centres account for 2% of global greenhouse-gas emissions; their power usage is expected to triple in the next decade[5]. And much financial and high-tech activity consists of coordinating resource extraction and manufacturing activities that have moved to other parts of the globe. Apple designs its iPhones in California, but 84% of the embodied carbon emissions of the phones come from their production in China, South Korea and other countries, mostly in Asia.

The low-carbon footprints prized by cities such as San Francisco and Seattle are little more than accounting tricks. The main method of carbon counting attributes to urban areas only the emissions resulting from in-city activities and regional power plants. Few studies count the full life cycle of emissions for all goods and services consumed by individuals and groups in cities, or emissions resulting from air travel. Those that do are telling. Consumption-based carbon counts for Shanghai, Seattle, San Francisco and London find more than double the per capita emissions of standard calculations. Almost 80% of San Franciscans' greenhouse-gas emissions, for example, are produced outside the city[6] (see 'Remote impacts').

The apparent low-carbon benefits of density fall dramatically when income and lifestyle are controlled for. Upper-income urban residents in the United States and Europe tend to consume more imported goods and services, fly more often, and drive out of the city more often than people living on lower incomes[7]. In the United Kingdom, during the explosion of low-cost air travel from the late 1980s to the early 2000s, the number of working-class passengers flying out of London increased by around 60%; wealthy passengers' trips increased by nearly 150%.

Although prosperous urban residents may commute by bicycle or public transport — the forms of low-carbon living most commonly cultivated by sustainability projects such as Freiburg's eco-neighbourhoods — their carbon footprints are enlarged greatly by their consumption practices and leisure travel. Economic activity and

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

urban density in post-industrial cities are inextricably linked with global networks of production, consumption and distribution.

KEY PLAYERS

It has become conventional wisdom that city leaders are more nimble and less ideological than their national counterparts. These two qualities, the story goes, allow leaders such as New York's former mayor Michael Bloomberg and former Bogotá mayor Enrique Peñalosa, along with networks such as the C40 Large Cities Leadership Group, to take the lead in confronting global sustainability challenges — even as international treaty efforts and national policymaking stall.

This 'urban turn' in policy and discourse captures important truths. But it obscures the fact that municipalities are more nimble because they wield less power. Municipal governments lack access to industrial policy, welfare systems and tax regimes. They have limited control over consumption patterns and large-scale infrastructure. And cities are bound by competitive pressures that pit them against each other in the pursuit of capital investment and talented workers. Municipalities thus tend to pursue sustainability policies that are also economic-development policies, and these disproportionately focus on affluent central business districts or residential areas designed to attract skilled professionals.

This challenges, for instance, the good intent of the United Nations' Sustainable Development Goals for cities. Reaching these goals requires strong national policy commitments to new regional infrastructure programmes, cash transfers to poor people, and local governance reform across urban regions.

State, provincial and national governments can apply sustainability policies across local jurisdictional lines. In the aftermath of Hurricane Sandy, which hit the US east coast in 2012, some of the dozens of small municipalities on the New Jersey Shore independently attempted to build new 'hard' seawalls, despite concerns that these would displace storm surges to their neighbours. Only higher levels of government can prevent such 'beggar-thy-neighbour' local politics.

And grass-roots groups bring about change from the bottom up. Community-based organizations, city-wide non-profit organizations and ad hoc social movements shape cities' built environment and lifestyle. But these groups are often overlooked in discussions about sustainability policy because most of them do not frame their work in environmental terms. They are more likely to speak of a broader 'right to the city'. Advocates for affordable housing and mass transit are proposing exactly the types of intervention that shrink individuals' carbon footprints and improve community resilience[8]. But they are rarely seen as prospective allies by green policymakers.

Sustainability efforts that are indifferent to concerns about affordability and that lack support from community members are less just and less likely to succeed. In New York City, an effort to implement a congestion charge in central Manhattan failed in the face of public opposition. New Yorkers in outer boroughs viewed the plan as elitist and indifferent to the concerns of poorer commuters. Still, some fledgling coalitions around equity and sustainability are emerging. Last year in São Paulo, a historic drought and state mismanagement of scarce water resources led housing movements and environmentalists — long at odds over how to deal with precarious waterside settlements — to come together around a common agenda of housing and water justice[9].

NEXT STEPS

First, urban environmental researchers need to supplement neighbourhood-specific and city-centric[10] measurements, such as walkability or commuting by public transport, with ones that better capture the broader dimensions of ecological sustainability and social equity. For instance, studies of changes to local transit systems should analyse the knock-on effects in regional housing and labour markets.

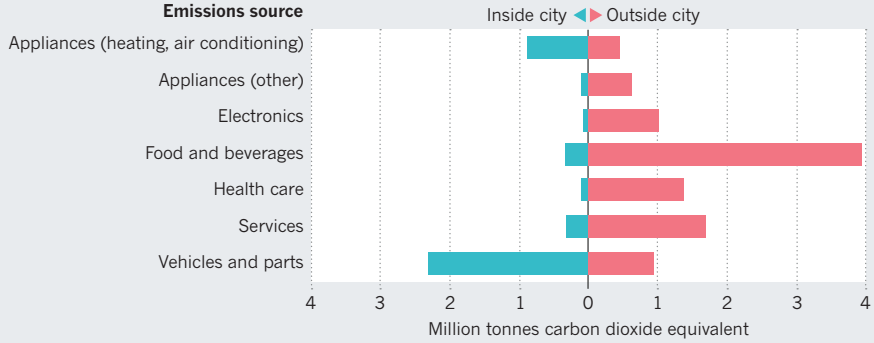
Second, multicity low-carbon policy networks such as the C40 and climate-focused organizations such as the World Resources Institute in Washington DC should insist on — and support — all large cities carrying out standardized, consumption-based carbon-footprint analyses. As well as providing more accurate accounts of specific cities' carbon footprints, this would underscore the extent to which emissions levels are correlated with class and income.

Third, policymakers should treat social equity and ecological effectiveness as mutually reinforcing dynamics in urban sustainability. They should bring the widest range of social movements to the table and see those groups' demands — such as revitalizing rent regulation and public housing — as central. This would entail more frequent meetings of larger groups of stakeholders and different metrics of policy success. But it would also yield more creative, sophisticated and encompassing policies that would have broader public support.

Only by expanding the spatial and social dimensions of urban policymaking can it be made truly sustainable and equitable. ■

REMOTE IMPACTS

In San Francisco, most of the carbon emissions associated with the consumption of goods by residents, firms and governments in 2008 arose beyond the city's limits — elsewhere in the United States or overseas. Yet municipal sustainability initiatives target only the metropolitan area.



SOURCE: REF. 6

David Wachsmuth *is assistant professor in the School of Urban Planning, McGill University, Montreal, Canada.* **Daniel Aldana Cohen** *is assistant professor in the Department of Sociology, University of Pennsylvania, Philadelphia, Pennsylvania, USA.* **Hillary Angelo** *is assistant professor in the Department of Sociology, University of California, Santa Cruz, Santa Cruz, California, USA.*
e-mail: david.wachsmuth@mcgill.ca

1. Checker, M. *City Soc.* **23,** 210–229 (2011).
2. Mössner, S. & Miller, B. *Regions Magazine* **300,** 18–20 (2015).
3. Anguelovski, I. *et al. J. Plan. Educ. Res.* http://dx.doi.org/10.1177/0739456X16645166 (2016).
4. Pellow, D. N. in *In the Nature of Cities: Urban Political Ecology and the Politics of Urban Metabolism* (eds Heynen, N. *et al.*) 226–244 (Routledge, 2006).
5. Bawden, T. 'Global warming: Data centres to consume three times as much energy in next decade, experts warn' *The Independent* (23 January 2016); available at http://go.nature.com/2aymx8h
6. Stanton, E. A. *et al. Consumption-Based Emissions Inventory For San Francisco* Version 2.0 (Stockholm Environment Institute-US Center, 2011); available at http://go.nature.com/2arawxw
7. Heinonen, J. *et al. Environ. Res. Lett.* **8,** 035050 (2013).
8. Ageyman, J., Bullard, R. D. & Evans, B. *Just Sustainabilities: Development in an Unequal World* (MIT Press, 2003).
9. Cohen, D. A. *Public Cult.* **28,** 261–289 (2016).
10. Angelo, H. & Wachsmuth, D. *Int. J. Urban Reg. Res.* **39,** 16–27 (2015).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



SELMUR/CINEMA REL. CIRO./THE KOBAL COLLECTION

Cliff Robertson as the title character in *Charly*, the 1968 film adaptation of *Flowers for Algernon*.

IN RETROSPECT

Flowers for Algernon

Ananyo Bhattacharya looks back at a science-fiction touchstone on the ethics of experimental biology.

By the time science-fiction writer Daniel Keyes died in 2014 at the age of 86, he had lived through vast upheavals in biomedical science, from the discovery of the DNA double helix to the sequencing of the human genome. But ethical oversight did not always keep pace. Keyes' novel *Flowers for Algernon*, 50 years old this year, highlights how often the need for oversight is ignored or flouted.

A case in point is a 1946–53 study conducted by Harvard University and the Massachusetts Institute of Technology, and sponsored in part by food conglomerate Quaker Oats. Dozens of boys with learning difficulties at the Walter E. Fernald State School in Waltham, Massachusetts, were fed cereals containing radioactive tracers to track how they absorbed iron and calcium. The boys were told only that they were joining a science club, and consent forms sent to their parents made no mention of radiation exposure. A US Department of Energy committee concluded in 1994 that it was "extremely unlikely" that the boys had been harmed by the radiation, but the disregard for their human rights is breathtaking. Other experiments, including some sanctioned by the US government, were much more egregious. Hundreds of African-American men involved in the Tuskegee Syphilis Study in Alabama from 1932 to 1972 were never told that they had the disease; nor were they treated, despite the availability of penicillin from the 1940s.

Flowers for Algernon
DANIEL KEYES
Harcourt, Brace & World: 1966.

The Tuskegee 'experiment' would never happen today, but the Massachusetts study's more subtle transgressions — in failing to fully regard the participants as ends in themselves, rather than a means to achieve the researchers' ends — remain relevant. It is this suppression of feeling for people and laboratory animals in the pursuit of scientific knowledge that Keyes captures in *Flowers for Algernon*.

Keyes' novel, based on a short story that he published in 1959, follows 32-year-old Charlie Gordon, who agrees to have an experimental brain operation that may help him to overcome his severe learning difficulties and increase his intelligence (he has an IQ of 68). The only subject to have previously undergone the procedure successfully is a lab mouse named Algernon. After the operation, Charlie's IQ rises rapidly; he soaks up new languages and knowledge of the arts and sciences. His journal entries, which make up the novel, chart his growing awareness of his own sexuality and emotions, particularly his feelings for his former teacher at the Beekman College Center for Retarded Adults.

More revealing of Keyes' intent is the evolving relationship between Charlie and Algernon. At first resentful of Algernon's superior intellect (the mouse easily beats him at navigating a maze), Charlie develops a strong bond with his fellow experimental subject. At the height of his genius, Charlie begins to investigate the experiment to advance the work. Soon realizing that it has flaws, he kidnaps Algernon to protect him. The regression that ends the book is so crushing that five publishers rejected the manuscript before it found a home. *Flowers for Algernon* became a best-seller (more than 5 million copies have been sold so far) and was adapted for the hit 1968 film *Charly*, starring Cliff Robertson. It still features in bioethics discussions.

Keyes had a degree in psychology and would later become a professor of creative writing at Ohio University in Athens. In between, he edited pulp magazine *Marvel Science Stories* and worked at Atlas Comics, the precursor to Marvel Comics. He also briefly taught English in New York City's public-school system. The empathy that suffuses the novel stems from his experience of teaching children with learning difficulties. When one student returned to classes after a long absence, Keyes noted that he had forgotten how to read. "He had lost it all," Keyes said. "It was a heartbreaker." His

NATURE.COM
For more on science in culture see:
nature.com/booksandarts

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

sympathy for Algernon seems to stem in part from dissecting a female mouse at university: Keyes was shaken when his incisions revealed "a cluster of tiny fetuses" in its uterus.

Despite his compassion for experimental subjects, human and animal, Keyes does not portray researchers as the evil geniuses of cultural cliché. Writing before modern ideas of informed consent were fully established in the late twentieth century, Keyes portrays the careerist psychologist Harold Nemur, who leads the trial, taking pains to get permission from Charlie's relatives to carry out the procedure. Neurosurgeon Jayson Strauss, who performs the operation, is concerned about Charlie's well-being throughout. What exercises Keyes is his scientists' failure to imagine Charlie as a whole human being before his intelligence-enhancing operation. Whereas Charlie's appreciation of Algernon's 'personhood' only grows, Nemur is unable to view Charlie as anything other than a sort of benign Frankenstein's monster.

"What exercises Keyes is his scientists' failure to imagine Charlie as a whole human being."

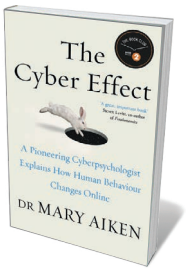
That hubris is sometimes evident today, when researchers fail to reflect fully on the consequences of their work (S. Aftergood *Nature* **536,** 271–272; 2016). A crop of findings suggests that the well-being of laboratory rodents has not been sufficiently prioritized. For example, mice are housed at around 20 °C, cooler than their preferred temperature of 30 °C (see *Nature* http://doi.org/bnh7; 2013). Many lab animals are also overweight. As well as being bad for their welfare, there is evidence that such conditions may skew experimental results (*Nature* **464,** 19; 2010).

This year, plans to make a synthetic human genome were criticized when discussions between more than 100 scientists took place behind closed doors and did not focus sufficiently on the proposal's ethical implications (*Nature* **534,** 163; 2016). Another controversy centred on the widely used HeLa cell line, derived in 1951 from the cervical tumour that killed an African American woman, Henrietta Lacks. But she had never consented to such use. In 2013, the cell-line genome was published — without permission from Lacks's living relatives.

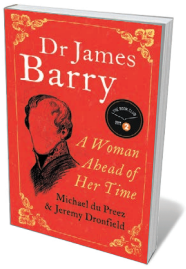
As the world enters the era of genome editing, it is tempting for scientists to monopolize the ethical debate once more. To avoid that temptation, researchers could do worse than turn to Keyes' astonishing *Flowers for Algernon*, a work that, tellingly, has never been out of print. ■

Ananyo Bhattacharya *is science correspondent at* The Economist *in London. e-mail: ananyobhattacharya@economist.com*

Books in brief



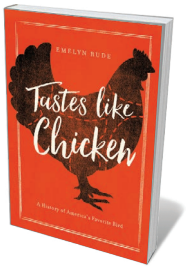
The Cyber Effect
Mary Aiken JOHN MURRAY *(2016)*
In this incisive tour of sociotechnology and its discontents, forensic cyberpsychologist Mary Aiken has much to say about children and the digital world. Parents addicted to mobile phones, for instance, fail to give babies the 'face time' they need to develop non-verbal communication skills; and the UK Association of Teachers and Lecturers has linked toddlers' tablet use with delays in speaking. With "compulsion loops" built into online games, and cybercommunities focused on extreme behaviours luring people in through online disinhibition, it's time for industrial accountability, she argues.



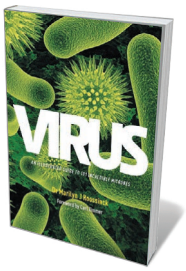
Dr James Barry: A Woman Ahead of Her Time
Michael du Preez and Jeremy Dronfield ONEWORLD *(2016)*
Over an illustrious career, Victorian surgeon James Barry became Britain's inspector-general of military hospitals, performed one of the first successful Caesarean sections in Africa and achieved the Crimean War's highest recovery rate. But under the overcoat, Barry was Margaret Ann Bulkley, who with the complicity of her mother and radical friends defied the rules and studied medicine at the University of Edinburgh. Urologist Michael du Preez and writer Jeremy Dronfield have drawn on fresh archive material for this nuanced biography of a medic with a mind-blowing secret.



Spare the Birds! George Bird Grinnell and the First Audubon Society
Carolyn Merchant YALE UNIVERSITY PRESS *(2016)*
From a fashion for feathers to habitat loss, US bird-life in the late nineteenth century faced pressing threats, prompting naturalist George Bird Grinnell — who had ties to the family of ornithologist John James Audubon — to launch a society and magazine in the great man's name. Carolyn Merchant's lavishly illustrated environmental history analyses Grinnell's contribution, from biographical writings on Audubon to delightful field descriptions of birds he portrayed — noting, for instance, how cedar waxwings (*Bombycilla cedrorum*) aid in reforestation by excreting undigested cherry stones.



Tastes Like Chicken: A History of America's Favorite Bird
Emelyn Rude PEGASUS *(2016)*
Andrew Lawler's 2014 *Why Did the Chicken Cross the World* (Atria; see *Nature* **515,** 490–491; 2014) gave us a natural and cultural history of *Gallus gallus domesticus*, from its south Asian origins to global ubiquity. In a breezy narrative brimming with retro recipes, culinary historian Emelyn Rude focuses on the history of US chicken consumption, currently 8.6 billion birds a year. From New York immigrants' foul "ornithological parks" of the 1880s and 1890s to the rise in global demand — which can push production at the expense of animal welfare — Rude reveals chicken as a troublesome taste.



Virus: An Illustrated Guide to 101 Incredible Microbes
Marilyn Roossinck IVY *(2016)*
Polio, Ebola, influenza — it's the viral villains that hit headlines, yet a number of viruses are benign. Environmental microbiologist Marilyn Roossinck sets the record straight with this stunning explication of 101 viruses that infect everything from humans to archaea. Along with basics on life cycles, transmission and more, Roossinck offers succinct descriptions, schematic drawings and a gallery of electron-microscopy images that have more than a passing resemblance to the paintings of Jackson Pollock and Wassily Kandinsky. **Barbara Kiser**

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Correspondence

Don't let climate crush coral efforts

June's International Coral Reef Symposium brought together more than 2,500 influential people who work on coral reefs, yet discussion centred on solving the global-scale issue of climate change, following a worldwide coral-bleaching event (*Nature* http://doi.org/bdmn; 2015). In our view, the symposium missed an important opportunity to develop real conservation outcomes for coral reefs at a local scale (see J. E. Cinner *et al. Nature* **535,** 416–419; 2016).

Discussions on climate change seem unproductive for environmental managers and scientists on the ground. Few individuals have a platform for engaging with global political leaders to drive the conservation agenda and influence policies that affect climate trajectories. Instead, we should be working together to develop strategies for local action that are robust to the uncertainty surrounding future climate scenarios.

We shall have to differentiate between those uncertainties that we can resolve at a local scale, such as the benefits of reducing overfishing or inputs of sediment and nutrients, and those that we cannot. To conserve coral reefs, we need objectives that can be turned into cost-efficient actions to deliver measurable, uncertainty-proof, local benefits.

Jennifer McGowan, Hugh P. Possingham *University of Queensland, St Lucia, Australia.*
Ken Anthony *Australian Institute of Marine Science, Townsville, Queensland, Australia.*
j.mcgowan@uq.edu.au

Social changes affect water quality too

Researchers need a better understanding of the effects of social shifts on river basins and water catchments, in addition to the impacts of climate change (A. Michalak *Nature* **535,** 349–350; 2016). To help safeguard water-catchment services against these social changes, communities should become more involved in water-management issues.

Urbanization and industrial and agricultural developments all generate changes in legislation, policy and demographics. These can adversely affect water-catchment services, which provide social, cultural and environmental benefits such as flood defence, recreational space, geodiversity and increased biodiversity.

Catchment-management initiatives, such as improving drinking-water quality and offsetting flood risk, are estimated to cost more than £100 billion (US$130 billion) over the next 15 years in England alone. By engaging with these initiatives, local communities can contribute to their implementation and ensure that they are cost-effective.

Alec Rolston *Dundalk Institute of Technology, Ireland.*
alec.rolston@dkit.ie

Stop marginalizing rare syndromes

Rare medical disorders are extremely challenging for patients and their families, and for researchers trying to study them. As a scientist and father of a 41-year-old son with CHARGE syndrome, which affects the heart, ears, eyes and other organs, I believe that we need stronger commitment, a more consistent approach and different types of knowledge and skills to move these investigations forward.

Promising initiatives include the International Rare Diseases Research Consortium and European Union projects on rare disorders under Horizon 2020. We still urgently need to improve and speed up genetic diagnosis, and to understand and mitigate the effects of these disorders on people's health.

This calls for an interdisciplinary strategy, but it faces significant methodological challenges. In a small study of 81 participants, for example, we found that those with Down's, Williams or Prader–Willi syndromes all had health problems related to diet and inactivity; also, there were important differences between disease groups and between women and men (M. Nordstrøm *et al. Food Nutr. Res.* **59,** 25487; 2015). Although such differences need to be underpinned by a wider evidence base from international collaborations, they indicate that a more sophisticated and personalized approach to care is paramount.

People with rare syndromes lack the political power of larger patient groups, so advances depend on holistic scientific insights and on funders overcoming their reluctance to support marginalized research.

Svein Olav Kolset *University of Oslo, Norway.*
s.o.kolset@medisin.uio.no

Venezuela's brain drain is accelerating

Your interview with the president of the Latin American Academy of Sciences, Claudio Bifano, barely reflects the scale of the scientific crisis in Venezuela (*Nature* **535,** 336–337; 2016). Its academic brain drain, for instance, is worse than indicated.

The figures you quote for scientists leaving Venezuela are from a preprint we released at the end of last year. Since then, the number has swollen rapidly from 1,504 to 1,820 — up from 1,783 in July, when the full paper was published (J. Requena and C. Caputo *Interciencia* **41,** 444–453; 2016). The latest tally represents almost 15% of Venezuela's scientists, who account for some 33% of its research publications. This rapid loss is coupled with the stalled recruitment of new talent.

This is in lamentable contrast to the end of the last century, when the research community grew by 200 or so Venezuelan scientists a year. Hugo Chávez took over as president in 1999 and, in my view, 16 years of disastrous science policies followed.

Jaime Requena *Academy of Physical, Mathematical and Natural Sciences, Venezuela.*
requena.j@gmail.com

Strengthen China's flood control

Heavy rainfall in China's Yangtze River basin as a result of the longest and strongest El Niño event for 65 years has led to severe flooding and economic losses of almost US$10 billion. Massive investments in flood defences after the 1998 deluge, which killed more than 4,000 people, proved inadequate. New tactics could help boost China's flood control.

Alongside better levees, enlarged reservoirs and improved early-warning systems, disaster-risk analysis can guide strategies for managing disasters (see S. L. Cutter *et al. Nature* **522,** 277–279; 2015). Disaster risk depends on the degree of hazard, exposure and vulnerability (Y. Zhou *et al. Risk Anal.* **34,** 614–639; 2014).

Accurate hazard assessment calls for a better understanding of extreme weather events and their rising frequency and intensity. Exposure calculations should factor in different population densities across the region. And vulnerability estimates should note the efficacy of early-warning systems and the resilience of local infrastructure (J. Birkmann *et al. Nat. Hazards* **67,** 193–211; 2013).

Speedy access to and effective dispersal of reliable disaster-risk information and of disaster-relief professionals will help to prevent and mitigate catastrophic outcomes, including secondary disasters such as landslides.

Yang Zhou, Yansui Liu *IGSNRR, Chinese Academy of Sciences (CAS); and Beijing Normal University, China.*
Wenxiang Wu *IGSNRR, CAS, Beijing, China.*
liuys@igsnrr.ac.cn

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Alfred G. Knudson

(1922–2016)

Cancer geneticist whose insights launched the search for tumour-suppressor genes.



COURTESY OF UT GRADUATE SCHOOL OF BIOMEDICAL SCIENCES AT HOUSTON

After years of observing children with the rare eye cancer retinoblastoma, Alfred Knudson proposed an explanation for how two different forms of it arise. His 'two-hit' hypothesis led to the realization that the loss of gene function, not just the activation of a cancer-causing gene, could cause cancer.

Knudson, who was born in Los Angeles, California, in 1922, died on 10 July, aged 93. After completing a bachelor of science degree at the California Institute of Technology (Caltech) in Pasadena in 1944, he earned a medical degree from Columbia University in New York City in 1947, and a PhD in biochemistry and genetics in 1956, also at Caltech. Knudson then spent years treating children in medical centres in California and New York.

During the 1950s and 1960s, cancer epidemiologists were preoccupied with trying to understand the environmental causes of the disease. In a 1953 paper, cancer biologist C. O. Nordling noted that in developed nations, the incidence of cancer seemed to increase with age (C. O. Nordling *Br. J. Cancer* **7,** 68–72; 1953). Nordling's proposal that the occurrence of cancer needed the accumulation of at least six sequential mutations was ultimately proved wrong. But his idea that cancer is caused by a certain number of 'hits' to the genome paved the way for Knudson's key insight.

Knudson had the foresight to focus on inherited tumours in childhood, which were relatively easy to study. The tumours could be counted and the early occurrence of the disease meant that there were fewer confounding factors to grapple with, such as the random genetic mutations that occur throughout life.

During his years in the clinic, Knudson had noticed that children with the hereditary form of retinoblastoma often developed multiple tumours in both eyes. By contrast, people with the 'sporadic' form developed a single tumour in only one eye. Also, in cases of hereditary retinoblastoma, the tumours typically occurred before the child was five; in sporadic cases, they occurred later in development.

On the basis of these observations, Knudson proposed that in hereditary retinoblastoma, one copy of the gene involved is mutated in the germ line (in reproductive cells such as eggs and sperm) and the other copy is mutated in somatic (non-reproductive) cells during the first few years of life, thus the cancer forms earlier. And because the germline mutation affects all somatic cells, these children are more prone to developing multiple tumours in both eyes. He argued that people who develop the other form are born with two normal alleles, both of which must become mutated in two 'sporadic' events in somatic cells, so the cancer develops later in life.

Knudson published his hypothesis in 1971 (A. G. Knudson *Proc. Natl Acad. Sci. USA* **68,** 820–823; 1971). He subsequently applied the same logic to other inherited tumours, such as Wilms' tumours (a type of kidney cancer) and those of the adrenal glands.

In 1983, cancer geneticist Webster Cavenee, then at the University of Utah in Salt Lake City, proposed that the genetic 'hits' in Knudson's mathematical models must be recessive, because the development of cancer happens only when both gene copies are mutated or lost. Using a technique called restriction fragment length polymorphism, Cavenee compared the DNA of tumours to that in normal tissues taken from people with retinoblastoma. He showed that the loss of heterozygosity (caused by a loss of the second, previously unaffected allele) led to cancer. (W. K. Cavenee *et al. Nature* **305,** 779–784; 1983).

Knudson's two-hit cancer hypothesis had a huge impact. Until this point, cancer was thought to be caused by the activation of oncogenes. Now the search was on for tumour-suppressor genes — whose loss of activity or function causes cancer. Knudson won the 1998 Albert Lasker Clinical Medical Research Award for his work on the genetic basis for cancer.

Knudson made other significant contributions through his leadership of one of the oldest cancer centres in the United States: the Fox Chase Cancer Center in Philadelphia, Pennsylvania. Joining in 1976, he spent 40 years there. He served as president (1980–82), scientific director (1982–83) and director of the centre's Institute for Cancer Research (1976–82).

The achievement of which he was most proud was giving Irwin Rose, a biochemist who joined the centre in 1963, US$50,000 so that Rose could extend the stay of two visiting scientists from Israel. Rose and these scientists, Avrum Hershko and Aaron Ciechanover, won the Nobel Prize in Chemistry in 2004 for discovering ubiquitin-mediated protein degradation. Cells use this process to break down and recycle protein; it has aided the development of several cancer drugs.

Alfred was a very supportive and approachable mentor, whose lack of patience for science that merely repeated the work of others kept everyone in his sphere striving for the new. He will be missed. ■

Carlo M. Croce *is distinguished professor and chair in the Department of Cancer Biology and Genetics, Ohio State University, Columbus, Ohio, USA. He met Alfred Knudson in the 1970s at the Wistar Institute in Philadelphia.*
e-mail: carlo.croce@osumc.edu

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



L. K. SANDERS, R. TRUBY, M. WEHNER, R. WOOD & J. LEWIS/HARVARD UNIV.

Figure 1 | The octobot. Wehner *et al.*[3] have made an octopus-shaped robot that is constructed completely from soft materials. The body houses a liquid-fuel supply and a fluidic system that controls a cyclic pattern of leg movements. Actuators that cause the legs to lift are visible as purple rectangles in the legs. Scale bar, 10 mm.

ROBOTICS

Generation soft

Meet the octobot, the first robot to be made entirely from soft materials. Powered by a chemical reaction and controlled by a fluidic logic circuit, it heralds a generation of soft robots that might surpass conventional machines. SEE LETTER P.451

BARBARA MAZZOLAI & VIRGILIO MATTOLI

Robots are typically used in manufacturing contexts that involve well-structured environments. These situations allow them to move following predefined procedures, limiting interactions with human operators for safety reasons. But if these machines were moved into 'real' environments outside factories, they would have to cope with uncertain situations, react and adapt to changing conditions, and interact safely with living organisms, including humans[1] — tough problems to solve using conventional technology made from hard materials. Robots made from soft, deformable materials[2] would be better able to grasp and manipulate unknown objects, and to move on unstructured and rough terrains, and might be less hazardous to people. On page 451, Wehner *et al.*[3] present the first robot that completely lacks rigid structures and control systems.

Soft body parts are important in many natural organisms. Animals such as squid, starfish and worms are composed almost entirely of soft materials and liquids, which increases their adaptability and robustness. There is therefore a growing belief that soft materials might help robotics technology to go beyond its current capabilities, by allowing robots to elongate, squeeze, climb and grow. For example, soft robotic arms inspired by octopuses can elongate[4], and soft robots that mimic caterpillars can roll and jump[5].

A notable attempt to develop a fully soft robot was reported[6] in 2011 by workers from the same research group as that of Wehner and colleagues. In that case, the robot itself was composed exclusively of soft materials, but a conventional pump-and-valve system was used to implement (actuate) different types of locomotion pneumatically, and was connected to the robot through cables. Wehner and colleagues now push the technological boundaries further, because not only are their robot's body and actuation units soft, but so also are the control system and power source, which are integrated into the robot. This makes it the first completely soft robot capable of operating without being tethered by cables.

The octopus-shaped robot — dubbed the 'octobot' by the authors (Fig. 1) — has eight arms moved by a pneumatic mechanism that relies on the expansion of embedded, inflatable compartments working as actuators. These actuators are integrated into a fluidic–pneumatic network powered by a liquid fuel (an aqueous solution of hydrogen peroxide). The fuel passes through reaction chambers that contain a platinum-based catalyst, which causes the hydrogen peroxide to decompose. This decomposition produces pressurized oxygen that inflates the actuators,

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

thus generating the arm movements.

Wehner *et al.* control the sequence of the octobot's arm movements using a completely soft fluidic circuit based on a system of valves that act as elements of logic gates. The circuit creates an oscillation that converts the inflow of pressurized fuel from the fuel storage chamber into outflows that alternate between different reaction chambers, until the system runs out of fuel. The octobot therefore repeats cycles of movements in which it first lifts four of its arms while lowering the other four, and then performs the reverse manoeuvre (see go.nature.com/2b3cn3s). The whole of the robot's body, including the fluidic circuit, is made of silicone-based materials that have different mechanical properties, tailored to the functional requirements of the various subsystems.

The realization of autonomous soft robots will require the integration of different materials and functionalities, such as actuation, powering and logic; the octobot represents the minimal system that demonstrates the potential of this approach. To achieve the required integration, Wehner and colleagues used a combination of advanced fabrication techniques — including micro-moulding[7], soft lithography[8] and multi-material embedded 3D printing[9] — to produce rubbery structures embedded with fluidic channels, spanning several orders of magnitude in length scales. Despite its apparent complexity, the customizability of this fabrication process allowed the authors to validate design modifications using a quick trial-and-error approach, so that the final device was rapidly optimized.

Wehner and colleagues' use of soft materials and continuum deformations — continuous bending of the arms to generate movement, rather than motion created by rigid structures connected by rotational joints — paves the way for further scientific and technological developments. The next steps are to develop computational control systems (such as more-sophisticated fluidic circuits) that allow a greater range of movement; to define new design rules for soft robots; and to adopt and improve manufacturing technologies.

Other challenges remain. For example, the forces that soft robots can exert on the environment might be limited, potentially restricting their applications. Moreover, the use of fluidic logic circuits as control systems, rather than conventional electronics, might limit the complexity of the behaviours that can be generated. A greater understanding of the properties of soft materials and how they interact with control systems and the environment is also needed, to produce desired robotic behaviour in real contexts[10].

Although soft robotics is still in its infancy, it holds great promise for several applications, such as servicing and inspecting machinery, search-and-rescue operations, and exploration. Soft robots might also open up new approaches to improving wellness and quality of life. Soft endoscopes that allow omnidirectional bending, elongation and tunable stiffening are already a reality[11], as are soft orthotic devices used for ankle and foot rehabilitation[12]. Wehner and colleagues' findings might help to guide research in these directions, contributing to the pillars of knowledge that will support the edifice of this new discipline. ■

Barbara Mazzolai *and* **Virgilio Mattoli** *are at the Center for Micro-BioRobotics, Istituto Italiano di Tecnologia, 56025 Pontedera, Pisa, Italy.*
e-mail: barbara.mazzolai@iit.it

1. Pfeifer, R., Lungarella, M. & Iida, F. *Science* **318,** 1088–1093 (2007).
2. Rus, D. & Tolley, M. T. *Nature* **521,** 467–475 (2015).
3. Wehner, M. *et al. Nature* **536,** 451–455 (2016).
4. Mazzolai, B., Margheri, L., Cianchetti, M., Dario, P. & Lasch, C. *Bioinspir. Biomim.* **7,** 025005 (2012).
5. Lin, H.-T., Leisk, G. G. & Trimmer, B. *Bioinspir. Biomim.* **6,** 026007 (2011).
6. Shepherd, R. F. *et al. Proc. Natl Acad. Sci. USA* **108,** 20400–20403 (2011).
7. Heckele, M. & Schomburg, W. K. *J. Micromech. Microeng.* **14,** R1–R14 (2004)
8. Qin, D., Xia, Y. & Whitesides, G. M. *Nature Protocols* **5,** 491–502 (2010).
9. Wu, W., DeConinck, A. & Lewis, J. A. *Adv. Mater.* **23,** H178–H183 (2011).
10. Paul, C. *Robot. Auton. Syst.* **54,** 619–630 (2006).
11. Ranzani, T., Gerboni, G., Cianchetti, M. & Menciassi, A. *Bioinspir. Biomim.* **10,** 035008 (2015).
12. Park, Y.-L. *et al. Bioinspir. Biomim.* **9,** 016007 (2014).

CANCER METABOLISM

Friendly neighbours feed tumour cells

In pancreatic cancer, neighbouring non-cancerous cells degrade their own proteins through a process called autophagy and release amino acids that are then taken up and used by the cancer cells. SEE LETTER P.479

JURRE J. KAMPHORST & EYAL GOTTLIEB

The cancer pancreatic ductal adenocarcinoma (PDAC) has a poor prognosis, and is an area of intense biomedical research. Previous studies[1,2] have shown that PDAC has a growth-promoting effect on pancreatic stellate cells in the surrounding connective tissue, and these cells reciprocate by supporting tumour growth and spread. On page 479, Sousa *et al.*[3] investigate the possibility that pancreatic stellate cells directly provide tumour cells with nutrients, in studies using mice and human pancreatic cells.

A key feature of PDAC is desmoplasia, in which the tumour becomes enmeshed and surrounded by dense, scar-like tissue[1,4]. As well as the tumour cells, desmoplasia involves other cell types, including pancreatic stellate cells, which are mainly responsible for the condition, and immune cells. Desmoplasia is a serious medical concern, because it forms a barrier to treatment with chemotherapeutic drugs as a result of poor blood perfusion into the tumour[5]. Paradoxically, desmoplasia is promoted by the release of signalling factors by the tumour, even though it also impedes the tumour's blood supply and hence its access to oxygen, glucose and other nutrients.

How can tumour cells obtain sufficient food and energy in the resulting nutrient-poor conditions? Sousa *et al.* showed that when PDAC cells were grown *in vitro* with either pancreatic stellate cells or even in the solution in which the stellate cells had been previously grown, oxygen consumption by the PDAC cells increased. This indicates that factors from the stellate cells can stimulate the activity of energy-producing organelles called mitochondria. Sousa and colleagues found that the increased mitochondrial activity is due to tumour consumption of the amino acid alanine, which is released at a high rate by the stellate cells.

The authors used alanine labelled with heavy carbon isotopes to trace the fate of this imported amino acid in tumour cells. Surprisingly, they found that alanine had an unusual metabolic fate in PDAC cells beyond its normal role in protein synthesis. When taken up by the cells, alanine can be metabolized to pyruvate in the aqueous intracellular region known as the cytosol. However, the authors found that instead it went primarily into the mitochondrial pool of metabolite molecules (Fig. 1).

Pyruvate is a key molecule in the mitochondrial tricarboxylic acid cycle, which generates cellular energy and feeds into biosynthetic pathways such as lipid synthesis. It is usually generated by glucose metabolism in a process called glycolysis. Another potential source of pyruvate is alanine transamination, in which pyruvate is produced by the removal of nitrogen from alanine. Alanine transamination can occur in the cytosol or the mitochondria.

Sousa *et al.* found that the alanine

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

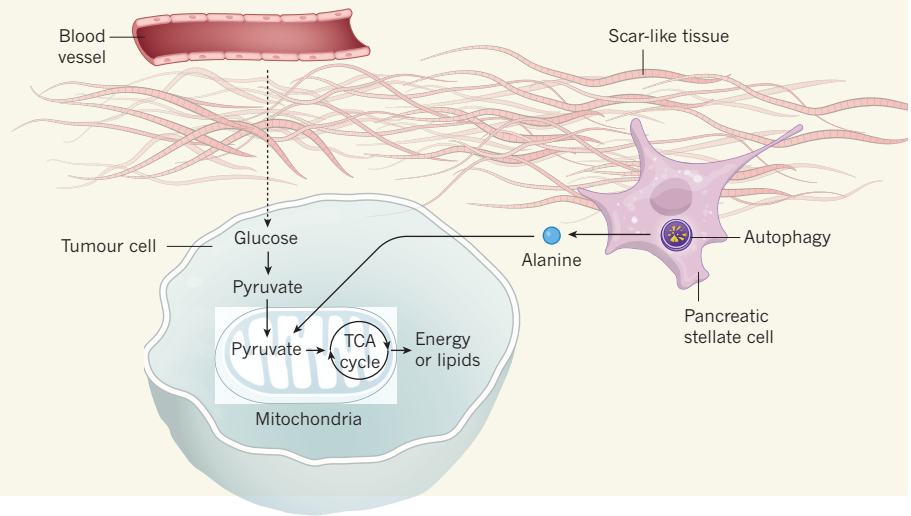


Figure 1 | Pancreatic stellate cells feed a tumour. Pancreatic ductal adenocarcinoma is a form of pancreatic cancer that is characterized by the build-up of dense, scar-like tissue pervading the tumour. This limits the delivery of oxygen and nutrients (including glucose) to the tumour cells by blood vessels. Sousa *et al.*[3] find that, near tumour cells, pancreatic stellate cells degrade their own proteins in a process called autophagy and release the resulting amino acids, most notably alanine, which are taken up and consumed by the tumour cells, facilitating cancer growth. In the tumour cell, alanine is metabolized to produce pyruvate in organelles called mitochondria, where pyruvate is used to produce energy and lipids through the tricarboxylic acid (TCA) cycle. The use of this pyruvate generated from alanine allows the limited glucose in the cell to be used for other metabolic purposes.

supplied to pancreatic cancer cells undergoes transamination primarily in the mitochondria, producing pyruvate that enters the tricarboxylic acid cycle to provide the cells with energy and lipid molecules. Such selective channelling of alanine to these biosynthetic pathways frees up glucose in the cell for use in other roles, such as production of the amino acid serine, which is required for nucleic-acid biosynthesis and hence cell growth and division.

Although Sousa and colleagues' work documents the metabolic support that alanine from pancreatic stellate cells can offer PDAC cells, it is unclear how and why alanine is channelled specifically to mitochondria. Alanine-derived pyruvate in the cytosol can also be transported to mitochondria for energy and biosynthesis purposes, and yet this takes place to a lesser degree. Direct transamination of alanine to pyruvate in the mitochondria seals its metabolic fate as a source of energy or biosynthetic molecules. One possible explanation for this compartmentalization of alanine transamination is the potentially higher availability of the nitrogen-acceptor molecule α-ketoglutarate in the mitochondria, which might enable fast and efficient production of pyruvate in response to the large supply of alanine from stellate cells.

Sousa *et al.* made a surprising finding — that a process called autophagy could affect a neighbouring cell. Autophagy is a survival mechanism used by cells to break down and metabolize their expendable proteins, lipids and other macromolecules when nutrients are scarce. It has been considered to be a process that acts only in the cell itself. However, Sousa and colleagues demonstrated that inhibiting autophagy in pancreatic stellate cells did not affect the cells' growth, but rather abolished alanine release and the resulting support of PDAC-cell growth both *in vitro* and *in vivo* (when PDAC cells and pancreatic stellate cells were transplanted together into mice).

The metabolic support of tumour cells by autophagy in pancreatic stellate cells provides a twist to the emerging tale of metabolic scavenging by PDAC cells. These latter cells have a high basal autophagy rate, helping them to survive metabolic hardship[6,7]. The previous observation[8] that PDAC-cell scavenging of extracellular proteins such as albumin can support PDAC metabolism and growth in nutrient-limited conditions highlights the tendency of these cancer cells to exploit diverse nutrient sources. Extracellular material is taken up by the cell through macropinocytosis, a cell-membrane-based transport system. These uptake and autophagy processes depend on organelles called lysosomes, and lysosomal activity was recently found[9] to be increased in PDAC.

Sousa and colleagues' work provides metabolic insight into the well-documented crosstalk between cancer cells and their neighbours. It has been demonstrated[10] that pancreatic cancer cells send signals to adjacent stellate cells that then reciprocate and alter the intracellular signalling and metabolism of the cancer cells. It will be interesting to learn which tumour-generated signals stimulate autophagy and alanine secretion from the pancreatic stellate cells. ■

Jurre J. Kamphorst *and* **Eyal Gottlieb** *are in the Cancer Metabolism Research Unit, Cancer Research UK Beatson Institute, Glasgow G61 1BD, UK.*
e-mails: jurre.kamphorst@glasgow.ac.uk; e.gottlieb@beatson.gla.ac.uk

1. Feig, C. *et al. Clin. Cancer Res.* **18,** 4266–4276 (2012).
2. Vonlaufen, A. *et al. Cancer Res.* **68,** 2085–2093 (2008).
3. Sousa, C. M. *et al. Nature* **536,** 479–483 (2016).
4. Sherman, M. H., Downes, M. & Evans, R. M. *Cancer Prevent. Res.* **5,** 3–10 (2012).
5. Olive, K. P. *et al. Science* **324,** 1457–1461 (2009).
6. Guo, J. Y. *et al. Genes Dev.* **25,** 460–470 (2011).
7. Yang, S. *et al. Genes Dev.* **25,** 717–729 (2011).
8. Commisso, C. *et al. Nature* **497,** 633–637 (2013).
9. Perera, R. M. *et al. Nature* **524,** 361–365 (2015).
10. Tape, C. J. *et al. Cell* **165,** 910–920 (2016).

This article was published online on 10 August 2016.

STRUCTURAL BIOLOGY

Signal locked in

A plant receptor protein interacts in an unusual way with the hormone it binds. The receptor cleaves the hormone, a fragment of which then binds covalently to the receptor and triggers a major receptor shape change. SEE LETTER P.469

KIMBERLEY C. SNOWDEN & BART J. JANSSEN

A receptor and its associated hormone are often thought of as a lock and key, in which the hormone key fits perfectly into the receptor lock, leading to a biological response — with the key ultimately being released intact from the receptor. Yao *et al.*[1] on page 469 and an accompanying paper in *Nature Chemical Biology* by de Saint Germain *et al.*[2] show that the reality is different for the plant hormones known as strigolactones. The strigolactone receptor protein cleaves strigolactone and forms a covalent bond between the receptor and a cleaved hormone fragment called the D-ring. This triggers a dramatic change in the shape of the receptor, exposing surfaces that can interact with signalling partners.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Strigolactones regulate aspects of plant development such as the growth of branches and the programmed death of leaves[3,4]. They are exuded by plants into the soil, where they stimulate interactions with symbiotic fungi that improve nutrient uptake. Strigolactones can also trigger the germination of parasitic plants, which can result in catastrophic crop failure, particularly in sub-Saharan Africa[3,4].

The evolutionarily conserved strigolactone receptor called AtD14 (also known as *Arabidopsis thaliana* DWARF14) is a member of the α/β hydrolase superfamily of proteins, characterized by a common core that coordinates three catalytic amino acids (serine, histidine and aspartate)[3–5]. This superfamily includes the enzyme acetylcholinesterase, one of the fastest known enzymes, which can catalyse 25,000 reactions per second; but it also includes an enzymatically inactive receptor for the plant hormone gibberellin[6,7]. The strigolactone receptor is unusual in that it acts both as a receptor and as an enzyme[5] that can cleave its own hormone. Even stranger, the rate of cleavage is very slow, a positively torpid rate of one reaction every 15–20 minutes[5].

Yao and colleagues[1] accomplished the difficult task of solving the X-ray crystal structure (at a near-atomic resolution of 3.3 ångströms) for AtD14 in a complex with the proteins D3 and ASK1, two components of an SCF complex involved in strigolactone signal propagation[3,4]. SCF complexes target proteins for degradation by another protein complex called the proteasome.

The structure revealed a fragment of the hormone covalently bound to two of the catalytic amino acids of AtD14. The authors' observations imply that, after the hormone strigolactone binds to AtD14, it is cleaved by a hydrolysis reaction catalysed by the receptor, and the D-ring hormone fragment becomes trapped in the catalytic active-site pocket, leading to a conformational change in AtD14 that promotes interaction with the SCF complex (Fig. 1).

The receptor in complex with the D-ring fragment and D3 undergoes a shape change when compared with structures of the receptor alone[5,8–11], with one of the four α-helix domains that comprise the receptor's 'lid' becoming extended in length, and another α-helix unfolding to form a loop. This change might explain previously observed destabilization of the receptor on hormone binding[5,11]. In addition, the catalytic active-site pocket is compressed in volume from 420 Å^3 to 80 Å^3, and the entrance to the pocket is closed. This observation indicates that release of the D-ring is unlikely without break-up of the complex, and could explain why the enzyme activity is so slow.

The greatest surprise from Yao and colleagues' crystal structure is the observation of a covalent bond between the D-ring and the receptor, because receptors don't usually bind their ligands in this manner. To confirm this finding, the authors used mass spectrometry to isolate a fragment of the receptor protein that showed a mass increase consistent with the D-ring bound to the histidine amino acid of the active site. This key observation is confirmed independently by de Saint Germain *et al.*[2] in the accompanying paper, which investigated the strigolactone receptor in pea plants (*Pisum sativum*). This group used fluorescent strigolactone analogues to show that the initial hydrolysis of these compounds was relatively fast, but that turnover to release the D-ring seemed to be extremely slow, because release was not observed within the experimental time period (up to 30 minutes).

Strigolactone promotes the degradation of the strigolactone receptor[12], but an unresolved question is whether the strigolactone D-ring covalently bound to the receptor is released before the receptor is destroyed. De Saint Germain and colleagues' observations suggest that there is no turnover of the D-ring to reset the receptor; however, *in vitro* experiments[3,5] show that, given sufficient time, the receptor can catalyse more than one reaction, which would be possible only if the D-ring is released. In addition, Yao *et al.* used mass spectrometry to detect release of the D-ring within an hour in *in vitro* assays. However, more studies, particularly *in vivo* experiments, are needed to clarify the kinetics of the reaction.

Yao *et al.* describe a plant with a mutant receptor that has increased enzymatic activity but is unlikely to undergo the conformational shift. The mutant plant does not respond to strigolactone (at least as far as branching is concerned). Meanwhile, de Saint Germain *et al.* show that some substrates that do not have biological activity can still be cleaved by the receptor; however, these substrates do not form the covalent linkage with the receptor because of the absence of a methyl group on the D-ring. Taken together, these findings suggest that the enzyme activity of the receptor is not sufficient for its biological function. Instead, the key function of the strigolactone receptor is the formation of the receptor–SCF complex, which is probably initiated by a conformational change in the receptor structure[3,4].

Yao *et al.* also show that a mutation in the strigolactone receptor that affects interaction with the SCF-complex protein D3 does not affect binding of the receptor to a downstream signalling component, which suggests that different receptor surfaces are involved in these interactions. X-ray crystal structures of the receptor bound to target proteins would clarify whether the receptor acts to present targets to D3 for subsequent degradation.

Although the studies by Yao and colleagues and de Saint Germain and colleagues answer some questions about this unusual receptor, they also raise new questions. We do not yet know the sequence in which the strigolactone receptor, SCF, and the proteins targeted for degradation are assembled into a complex. Nor is it known whether receptor enzyme activity is fast or slow *in vivo*, and more work is needed to understand what the active form of the receptor–D-ring complex is during signalling. In particular, does it involve linkage

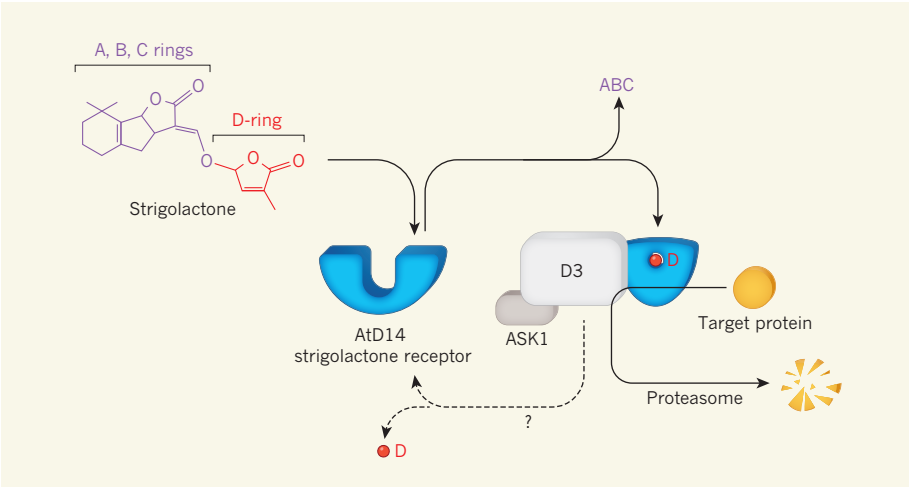


Figure 1 | An unusual hormone–receptor interaction. In the absence of the hormone strigolactone, the AtD14 strigolactone receptor protein has a large, open cavity that contains a catalytic site. This binds and hydrolyses strigolactone, then releases the ABC-ringed portion of the hormone, and it has been found that the receptor traps the hormone D-ring as a covalently bound intermediate. Yao *et al.*[1] identified the trapped D-ring intermediate in X-ray crystal structures of AtD14 from the plant *Arabidopsis thaliana*, and de Saint Germain *et al.*[2] monitored the hormone interaction with the strigolactone receptor of pea plants using fluorescent substrate analogues. AtD14 subsequently undergoes a conformational shift that allows the protein D3 to bind the receptor. The proteins D3 and ASK1 form part of the SCF complex, which targets proteins for degradation. The receptor–SCF complex adds a molecular tag (not shown) to targeted signalling proteins, resulting in their degradation by the proteasome apparatus. If the D-ring were released from the receptor, this would reset the receptor to its original state, but whether the D-ring is released *in vivo* is not known.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

of the D-ring to just the histidine or to both the histidine and serine catalytic amino acids? We also do not know the extent of conformational change in the receptor before D3 binds.

A broader question is how does the structural diversity of the many known strigolactones fit into the varied roles of these compounds in signalling within and between plants and symbiotic fungi? Has this mode of receptor action evolved to cope with the variation in strigolactones necessary in the arms race to evade parasitic plants? Only time will tell if this mode of receptor–ligand action is used by other receptors. ■

Kimberley C. Snowden *and* **Bart J. Janssen** *are in the Breeding and Genomics Group, New Zealand Institute for Plant and Food Research, Auckland 1025, New Zealand.*
e-mails: kimberley.snowden@plantandfood.co.nz; bart.janssen@plantandfood.co.nz

1. Yao, R. *et al. Nature* **536,** 469–473 (2016).
2. de Saint Germain, A. *et al. Nature Chem. Biol.* http://dx.doi.org/10.1038/nchembio.2147 (2016).
3. Al-Babili, S. & Bouwmeester, H. J. *Annu. Rev. Plant Biol.* **66,** 161–186 (2015).
4. Khosla, A. & Nelson, D. C. *Curr. Opin. Plant Biol.* **33,** 57–63 (2016).
5. Hamiaux, C. *et al. Curr. Biol.* **22,** 2032–2036 (2012).
6. Holmquist, M. *Curr. Protein Pept. Sci.* **1,** 209–235 (2000).
7. Ueguchi-Tanaka, M. *et al. Nature* **437,** 693–698 (2005).
8. Zhao, L.-H. *et al. Cell Res.* **23,** 436–439 (2013).
9. Nakamura, H. *et al. Nature Commun.* **4,** 2613 (2013).
10. Kagiyama, M. *et al. Genes Cells* **18,** 147–160 (2013).
11. Zhao, L. H. *et al. Cell Res.* **25,** 1219–1236 (2015).
12. Chevalier, F. *et al. Plant Cell* **26,** 1134–1150 (2014).

This article was published online on 1 August 2016.

INORGANIC CHEMISTRY

Fleeting glimpse of an elusive element

A heroic effort to characterize the chemistry of actinium, a short-lived radioactive element, reveals surprising differences in behaviour compared with other elements in the actinide series.

THOMAS E. ALBRECHT-SCHMITT

It is now possible to study the chemistry of radioactive elements that have short half-lives using state-of-the-art techniques that were not available when the elements were first discovered. Writing in *Nature Communications*, Ferrier *et al.*[1] report just such an investigation of the poorly understood heavy element actinium (Ac) — the first element in the actinide series of the periodic table. They reveal that the chemical behaviour of actinium in water differs from that of the heavier actinides, and in so doing provide a clue that might enable actinium to be used in cancer radiotherapy.

Like all elements beyond bismuth in the periodic table, actinium has no stable isotopes. Moreover, the only available isotopes, ^{225}Ac and ^{227}Ac, must be extracted from the decay of other radioactive elements. The parents of these actinium isotopes include several isotopes of radium — which is not readily available and also decays, to form extremely dangerous radon gas — as well as rare isotopes of thorium, uranium and protactinium. As a result, the supply of actinium is typically limited to only a few micrograms at a time.

The challenges continue to mount in the case of ^{225}Ac, because it has a half-life of just 10 days. To make matters worse, the chemistry of actinium cannot be interrogated using standard spectroscopic techniques, because its only oxidation state is +3, and the associated ion, Ac^{3+}, has a 'closed-shell' electron configuration — rendering it invisible to those techniques. The combination of intense radioactivity and spectroscopic invisibility makes efforts to tackle actinium chemistry nothing short of heroic.

Understanding the fundamental chemistry and physical properties of elusive elements is a worthy enterprise in itself, but there is another reason for the interest in actinium: ^{225}Ac has been recognized[2] as a promising candidate for treating cancer through a process called targeted alpha therapy. This therapy relies on the fact that ^{225}Ac decays by emitting α-particles — which consist of two protons and two neutrons, and are one of the most common forms of radiation emitted by decaying radioactive elements. These particles pack much more energy than the other types of radiation commonly formed during radioactive decay (β-particles and γ-rays), but they can usually travel through only the first few layers of cells in tissue[3]. This makes them ideal for cancer therapy, because it means they can destroy cancer cells without damaging the surrounding tissue.

Understanding the fundamental chemistry and physical properties of elusive elements is a worthy enterprise.

However, to develop radioisotopes for this kind of treatment, molecules that trap the isotope's ion and deliver it to specific areas of the body are needed. Such molecules are called chelating agents, and must have remarkably high binding specificity for the radioisotope being delivered — otherwise, the isotope could be displaced from the chelating agent by other, biologically available metals, such as iron, when the isotope enters the bloodstream. Herein lies the rub: because the chemistry of actinium is sufficiently different from that of other elements with +3 oxidation states (trivalent elements), we do not have appropriate chelating agents for actinium. Moreover, we cannot design them without knowing how molecules and ions bind to Ac^{3+} ions in water. The goal of Ferrier and colleagues' work was to uncover why actinium's chemistry deviates from that of other trivalent elements in the lanthanides and actinides, the two series of elements with which actinium is grouped in the periodic table.

Because many actinides are scarce and radioactive, molecules that probe their reactivity must be developed using analogous, non-radioactive elements. This often takes years of effort. Fortunately, trivalent actinides can have chemistry similar to that of trivalent lanthanides, of which only promethium is radioactive. More specifically, a lanthanide ion can typically be found that has a similar ionic radius to that of a target actinide, and can therefore act as the actinide's substitute during efforts to explore chemical reactivity, preventing the depletion of precious radionuclide resources. Even when the origin of the differences between actinides and lanthanides is being studied, lanthanides provide a good first approximation of the chemical behaviour of actinides.

However, finding a lanthanide ion of similar size to Ac^{3+} presents a challenge, because the ionic radius of actinium, although not established with a high degree of precision, must be much larger than that of lanthanum (La, the largest lanthanide). In fact, the chemistry of La^{3+} might well lead researchers astray, because it is known to be unusual — lanthanum is the only element to form ions that commonly interact directly with ten molecules[4] (to use the jargon, it often has ten ligand molecules in its inner sphere).

So how did Ferrier *et al.* overcome the practical issues of working with actinium? They took advantage of recent upgrades at the

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

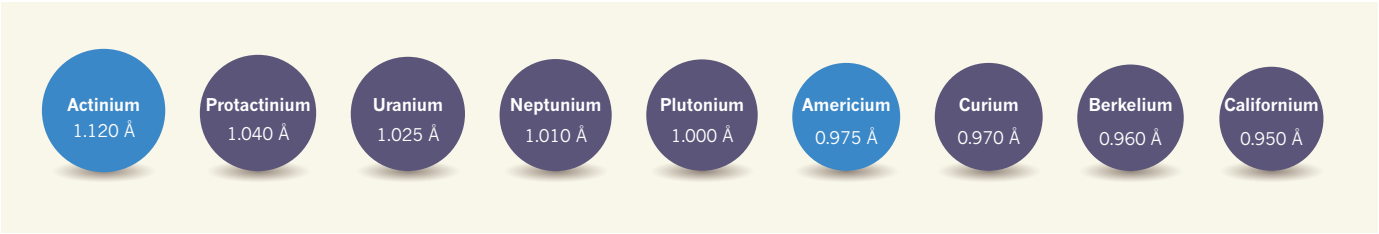


Figure 1 | Actinide contraction. The radii of the tripositive ions of the first elements of the periodic table's actinide series become progressively smaller[3]. (Ionic radii are given in ångströms, and ions are shown to scale; thorium, the second element in the series, is not shown because it does not form a stable tripositive ion in water.) The size of each ion correlates with the ion's ability to bind to certain kinds of ligand molecule. Ferrier *et al.*[1] report that the ligand-binding behaviour of actinium differs substantially from that of americium.

Stanford Synchrotron Radiation Lightsource, a facility in California capable of producing extremely bright X-rays, to study microgram samples of actinium using a technique called X-ray absorption fine-structure spectroscopy — which actinium ions are not invisible to. More specifically, the authors studied solutions of an Ac^{3+} salt dissolved in hydrochloric acid to observe how water molecules and chloride ions bind to the Ac^{3+} ions. They also used computational modelling to help interpret the results.

The authors find that Ac^{3+} binds to nine ligands in aqueous solutions, as is common for early actinides (actinium to americium in the actinide series). This was perhaps unexpected: given the large ionic radius of Ac^{3+}, one might have anticipated that its inner sphere of ligands would have been larger. But the real surprise was the finding that the inner sphere incorporates more chloride ions than was previously thought: the dominant species observed was $Ac(H_2O)_6Cl_3$. For comparison, Ferrier *et al.* studied solutions of americium (Am), because americium is the first trivalent actinide after actinium not to readily undergo changes of oxidation state. The primary species they observed in hydrochloric acid was $[Am(H_2O)_8Cl]^{2+}$.

One of the few predictable trends known for the actinides is that the ionic radius of the tripositive ions contracts as one traverses the series from actinium to lawrencium. The contraction is quite consistent, averaging about 0.01 ångströms between neighbouring actinides[5] (Fig. 1). Although the contraction is gradual, the chemistry of the ions often changes considerably in tandem with each contraction, so that adjacent actinide compounds exhibit quite different atomic structures, reactivities and electronic properties. Protactinium[6], plutonium[7] and californium[8] are all good examples of actinides whose ion chemistry differs substantially from that of their neighbours in the periodic table.

The contracting ionic radius also means that the later actinides in the series have higher charge densities than earlier ones. Actinide ions in general are said to be 'hard' cations, which means that they bind preferentially to small ligands such as water and fluoride ions. But Ac^{3+} should be the 'softest' of the trivalent actinides, binding preferentially to softer, larger ligands. Ferrier and colleagues' observation that actinium binds to a larger number of chloride ions in water than americium is consistent with this theory, because chloride ions are larger than water molecules. This knowledge can be used to design chelating agents for actinium-based radiopharmaceuticals: such agents should contain more soft atoms than are typically used in these compounds. ■

Thomas E. Albrecht-Schmitt *is in the Department of Chemistry & Biochemistry, Florida State University, Tallahassee, Florida 32306, USA.*
e-mail: albrecht-schmitt@chem.fsu.edu

1. Ferrier, M. G. *et al. Nature Commun.* http://dx.doi.org/10.1038/ncomms12312 (2016).
2. Ulmert, D., Abou, D., Lilja, H., Larson, S. & Thorek, D. *J. Nucl. Med.* **56** (suppl. 3), 281 (2015).
3. Zalutsky, M. R. & Vaidyanathan, G. *Curr. Pharm. Design* **6,** 1433–1455 (2000).
4. Ruis-Martinez, A. & Alvarez, S. *Chem. Eur. J.* **15,** 7470–7480 (2009).
5. Burns, J. H., Peterson, J. R. & Baybarz, R. D. *J. Inorg. Nucl. Chem.* **35,** 1171–1177 (1973).
6. De Sio, S. M. & Wilson, R. E. *Inorg. Chem.* **53,** 12643–12649 (2014).
7. Janoschek, M. *et al. Sci. Adv.* **1,** e1500188 (2015).
8. Polinski, M. J. *et al. Nature Chem.* **6,** 387–392 (2014).

NEUROSCIENCE

Memories linked within a window of time

In mice, two fear-associated memories that are created close in time are represented in the brain's amygdala by the activation of overlapping ensembles of neurons. As a result, eliminating the fear of one memory also extinguishes fear of the other.

HOWARD EICHENBAUM

Two of my strongest memories are of emotional but unrelated events that occurred on the same day. On 20 July 1969, I successfully finished a gruelling month-long, round-the-clock experiment. That evening, I saw men first walk on the Moon. These two events are, for me, forever connected. Why? Writing in *Science*, Rashid *et al.*[1] offer an explanation — emotionally charged events that occur close in time are bound together because of an overlap between the ensembles of neurons that are excited when the memories are laid down.

In their study, Rashid and colleagues trained mice to fear a particular tone (tone 1) by pairing the sound with a mild electric shock. Then, after an interval of between 1.5 and 24 hours, they trained the same animals to fear a different tone (tone 2). If the authors subsequently extinguished the animals' fear of tone 2, by repeatedly using it without shock, fear responses to tone 1 also decreased — but only if the initial training events had occurred within 6 hours of each other. Thus, a selective link forms between memories of fear-associated events that occur close together.

A brain region called the amygdala is essential for associating cues with shock. Select ensembles of neurons in this region are activated in mice during fear-associated training, and are then reactivated when the animal recalls the event. Using a sophisticated method for marking neurons activated during each event, Rashid *et al.* found a high degree of overlap (co-allocation) between the neurons activated by the two memories if the memories were created within 6 hours of each other, but not when created 24 hours apart (Fig. 1).

Next, the authors used elegant genetic tools to force activation of a common set of amygdala neurons during both learning events — a trick that artificially linked memories

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



50 Years Ago

The Meteorological Office is cheerful if restrained about the success of its long range weather prediction service in its annual report for 1965 … Long range forecasts for a period of thirty days ahead have been published twice a month for the past three years … the Meteorological Office says that results have been slightly better than expected … forecasts are assessed after the event … and "marks are given for the accuracy of forecasts of temperature, rainfall and additional information". Predicting temperature seems to be the easiest, with twenty-eight out of fifty forecasts ranking for the mark "good agreement". On rainfall, however, agreement between forecasts and reality was good on fourteen occasions, moderate on seventeen and deserving the description "little agreement" on nineteen occasions … The long range forecasts are based on searches of records going back to the middle of the nineteenth century for analogous patterns of mean temperature and mean pressure in the northern hemisphere … What might be called objective long range forecasting is reckoned to be an extremely distant prospect.
From *Nature* 20 August 1966

100 Years Ago

The firing of very heavy guns at a great distance was clearly audible at Harpenden throughout the days of August 7 and 8, as well as on previous occasions. The direction of the sound is evidently from the south-east, and that of each explosion lasts about two seconds. Our elevation is 440ft., and the local wind has been from west to north-west. The distance between Harpenden and Bapaume would be about 185 miles.
From *Nature* 17 August 1916

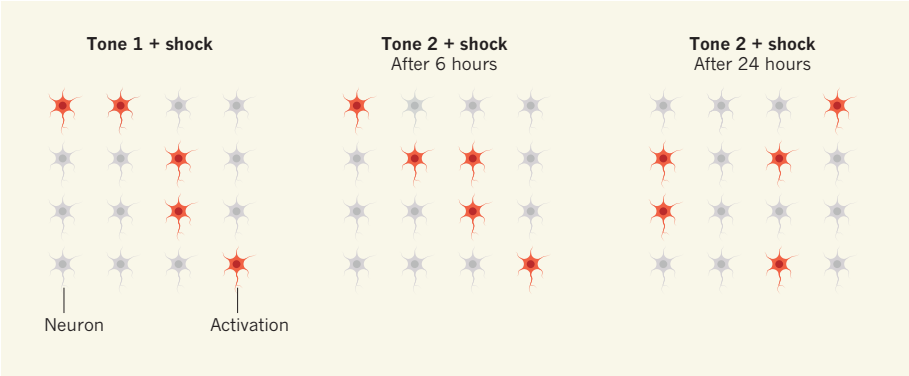


Figure 1 | Co-allocating cell populations to memories by time. Rashid *et al.*[1] trained mice to fear a tone (tone 1) that was coupled to a mild electric shock. They then repeated the training for a different tone (tone 2), either 6 or 24 hours later. Overlapping populations of neurons in the brain's amygdala were excited during the training sessions 6 hours apart, but not those 24 hours apart. As a result, the memories of the training sessions that occurred close together became linked.

that occurred outside the 6-hour window. By contrast, when they inhibited excitation of the same population, preventing co-allocation of neuronal ensembles, memories that occurred within the 6-hour window were separated. These experiments confirm that common excitability of neurons in the amygdala is responsible for co-allocating neurons to memories within a window of time.

A related paper by Cai *et al.*[2] was recently published in *Nature*, and reported a similar finding — spatial memories that are acquired near in time are associated with overlapping neuronal ensembles in the brain's hippocampus. In this study, the authors monitored neuronal activity in mice exposed to one environmental context (context A), then a week later to two other contexts (contexts B and C) separated by five hours. In line with Rashid and colleagues' findings, Cai *et al.* observed co-allocation of activated hippocampal neurons as animals explored contexts five hours apart, but not seven days apart. They then demonstrated the behavioural significance of this co-allocation by showing that subsequent pairing of shock with context C resulted in fear of that context and of context B, but not of context A.

Importantly, the memories of contexts B and C were distinct — subsequently extinguishing the fear of context C did not extinguish the fear of context B. This is in contrast to Rashid and colleagues' findings, and suggests a distinction in how timing contributes to the integration of memories in different systems. The results in the amygdala reflect the region's role in generalizing defensive reactions to aversive stimuli. By contrast, in the hippocampus, the linkage of distinct memories promotes recall of one memory when cued by another, reflecting the role of this region in supporting the elaborate organizational structure and flexibility of our conscious memories.

The role of time in binding memories has been recognized since Aristotle's day[3]. Our current understanding of this process is embodied in a theory called temporal context memory (TCM), which was originally developed to explain why people have a better memory for words that occur close together in a list than for those farther apart. The theory was extended in 2013 to explain why people better remember events that occur within a few thousand seconds of each other[4]. TCM theory posits that the transient overlapping of neural-activity patterns creates a gradually changing activity landscape that can link memories that are created close together in time and can separate memories that are created far apart. The present results confirm TCM theory.

Consistent with TCM, a neurophysiological study in rats has shown[5] that firing patterns of neural ensembles in the hippocampus gradually change over time. This change enables the animals to remember the order in which events occurred during an experience. Moreover, the gradual evolution of the neural networks that represent experiences continues for days[6,7]. Brain-imaging studies in humans have also documented a gradual evolution of hippocampal neuronal activity associated with successful recall of the temporal order of events and with distinguishing events that occur at different times[8,9]. Thus, in addition to the role of time in linking and separating memories described in the current studies, time can also sequentially organize memories.

Of course, time is only one variable that can link and separate memory representations. Previous studies have demonstrated that other types of contextual manipulation, such as shared or distinct stimuli[9] and consistent or opposing reward associations for the same stimulus[10], can link or separate neural representations of memories that occur close in time. Nonetheless, Rashid *et al.* and Cai *et al.* extend our understanding of the role of time in integrating and separating emotional and spatial memories, and reveal that allocation of neural ensembles underlies the powerful role of evolving temporal contexts in linking and separating memories. ■

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Howard Eichenbaum *is in the Center for Memory and Brain, Boston University, Boston, Massachusetts 02215, USA.*
e-mail: hbe@bu.edu

1. Rashid, A. J. *et al. Science* **353,** 383–387 (2016).
2. Cai, D. J. *et al. Nature* **534,** 115–118 (2016).
3. Aristotle. 'On memory and reminiscence.' http://classics.mit.edu/Aristotle/memory.html (transl. Beare, J. I.) from *The Works of Aristotle* Vol. 3 (ed. Ross, W. D.) (Clarendon, 1930).
4. Howard, M. W. & Eichenbaum, H. *J. Exp. Psychol. Gen.* **142,** 1211–1230 (2013).
5. Manns, J. R., Howard, M. W. & Eichenbaum, H. *Neuron* **56,** 530–540 (2007).
6. Mankin, E. A. *et al. Proc. Natl Acad. Sci. USA* **109,** 19462–19467 (2012).
7. Ziv, Y. *et al. Nature Neurosci.* **16,** 264–266 (2013).
8. Hsieh, L.-T., Gruber, M. J., Jenkins, L. J. & Ranganath, C. *Neuron* **81,** 1165-1178 (2014).
9. Ezzyat, Y. & Davachi, L. *Neuron* **81,** 1179–1189 (2014).
10. McKenzie, S. *et al. Neuron* **83,** 202–215 (2014).

CELL DIVISION

Mitotic regulation comes into focus

Structural studies provide insight into the mechanisms governing a checkpoint in cell division that prevents chromosomes from segregating before they are properly aligned on a structure called the mitotic spindle. SEE ARTICLE P.431

DAVID O. MORGAN

Late in cell division, duplicated chromosomes are pulled apart by a structure called the mitotic spindle. This process, called mitosis, is beautiful to watch, but fraught with danger: errors can produce daughter cells that have unequal genomes, setting the stage for cell death or cancer. The cell goes to great lengths to avoid these mistakes, and uppermost among its safety mechanisms is the spindle assembly checkpoint[1] (SAC) — a regulatory system that prevents the cell from attempting to separate chromosomes if they are not properly attached to the spindle. The key protein components of the SAC were first reported in the summer of 1991 (refs 2,3). Precisely 25 years on, two papers (Alfieri *et al.*[4] on page 431 and Yamaguchi *et al.*[5] in *Molecular Cell*) describe in near-atomic detail how the SAC prevents chromosome separation.

The central protagonist in this story is the anaphase-promoting complex/cyclosome (APC/C)[6]. This enzyme attaches the protein ubiquitin to specific substrate proteins, tagging them for destruction and thereby unleashing chromosome separation and the completion of mitosis. A key APC/C target, for instance, is the protein securin, which inhibits a protease enzyme that cuts the proteins holding duplicated chromosomes together. Substrate binding to the APC/C depends on a protein called Cdc20 (Fig. 1a), which binds the APC/C during mitosis and recruits substrates through interactions with short linear amino-acid sequences in the targets called degrons.

How does the SAC block chromosome separation? Chromosomes that are not correctly attached to the spindle produce a mitotic checkpoint complex (MCC), which contains three SAC components (Mad2, BubR1 and Bub3) and Cdc20. The MCC binds and inhibits Cdc20-bound APC/C (APC/C^{Cdc20}), forming a large complex called APC/C^{MCC} that contains two copies of Cdc20 (refs 1,6,7). Alfieri *et al.* and Yamaguchi *et al.* use cryo-electron microscopy to unveil high-resolution structures of this giant assembly.

A structure of this size and complexity, backed up by 25 years of genetic, biochemical and structural analysis, contains an overwhelming amount of information for the aficionado, but its chief value lies in the precise description of the mechanisms by which the MCC inhibits APC/C^{Cdc20}. Both studies show that the MCC interacts with the front face of the APC/C, next to the pre-bound Cdc20 subunit (Fig. 1b). The BubR1 subunit acts as a pseudosubstrate inhibitor — it contains two copies of each of the three major degron sequences, and wraps around the two Cdc20 subunits to occupy all degron-binding sites on both, thereby blocking substrate binding to the APC/C. It is hard to find a more striking illustration of the power of short linear motifs such as degrons in cell regulation. In addition, the MCC prevents binding between the APC/C and the E2 coenzyme, which normally provides ubiquitin for transfer to target proteins.

The new structures also offer a potential explanation for previous evidence that some proteins are ubiquitinated by APC/C^{Cdc20} even when the complex is bound to the MCC. For example, the APC/C substrate proteins cyclin A and Nek2A are degraded early in mitosis, when the SAC is active[6]. Similarly, the Cdc20 subunit of the MCC is tagged with ubiquitin by the APC/C, promoting MCC turnover[1]. How does APC/C^{MCC} modify these proteins in spite of the mechanisms suppressing its activity?

In addition to the 'closed' conformation described above, the two studies reveal a less-common 'open' state in which the MCC is shifted to allow binding of E2. This transient state enables ubiquitination of Cdc20

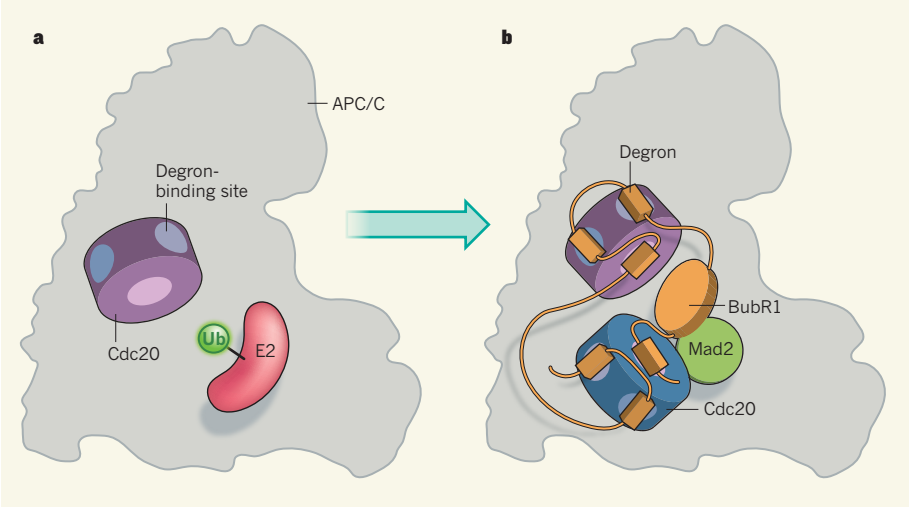


Figure 1 | Deciphering inhibition of cell division. a, The enzyme anaphase-promoting complex/cyclosome (APC/C), depends on an activator subunit called Cdc20 for activity and substrate binding. Cdc20 contains binding sites for three short amino-acid sequences called degrons. APC/C targets (not shown) that contain these degrons bind these sites and are tagged with the protein ubiquitin (Ub), which is transferred from an E2 coenzyme bound nearby. **b**, Alfieri *et al.*[4] and Yamaguchi *et al.*[5] report that the mitotic checkpoint complex, which comprises the proteins BubR1, Bub3, Mad2 and another Cdc20, interacts with Cdc20-bound APC/C to form a closed inhibitory state in which E2 binding is blocked (Bub3 is not shown, because its position could not be defined in the groups' structures). Disordered regions in the BubR1 subunit contain degrons that occupy all six degron-binding sites on the two Cdc20 molecules, preventing substrate binding.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

by APC/C^{MCC}, and might also explain the ubiquitination of cyclin A and Nek2A, which bind APC/C^{Cdc20} not only at sites blocked by BubR1, but also at other sites (ref. 6). With the open structure in hand, we are in a good position to unravel these mechanisms. It will also be important to explore the dynamics of the shift between conformational states, and how other regulatory proteins influence these dynamics.

These APC/C^{MCC} structures come close on the heels of a structural study[8] by Zhang *et al.* that addressed another major question in APC/C regulation. It is known[9] that the mitotic protein-kinase enzyme Cdk1–cyclin B phosphorylates the APC/C to promote its activation by Cdc20. APC/C^{Cdc20} then triggers cyclin B degradation to inactivate Cdk1. This negative feedback is thought to be the basis for the oscillator that drives the rise and fall of Cdk1 activity during the cell cycle[9], but much about this scheme has been unclear. How does phosphorylation activate APC/C^{Cdc20}? How is APC/C activation delayed to allow Cdk1 activity to rise in early mitosis — even in cell types in which the SAC is not present?

Zhang and colleagues' analysis, together with recent biochemical studies[10,11], yields valuable clues. Phosphorylation of a loop region in the Apc3 subunit of the APC/C provides docking sites for a phosphate-binding subunit of Cdk1–cyclin B. Docking of Cdk1–cyclin B enhances its activity towards suboptimal phosphorylation sites in a loop on a different APC/C subunit, Apc1. This disordered loop contains a short 'autoinhibitory' segment that occupies a Cdc20-binding site on the APC/C, but phosphorylation displaces the segment, allowing Cdc20 to bind and activate the APC/C. The slow, multistep nature of this process provides a plausible mechanism for introducing a delay between Cdk1 activation and APC/C^{Cdc20} activation, as is required for negative feedback to produce a robust oscillator[9,12].

We have not heard the last word on APC/C phosphoregulation. The APC/C contains many phosphorylation sites in addition to those described above[8,10,11], leaving open the possibility of other regulatory mechanisms, or connections between phosphorylation and the SAC.

Another issue also remains unresolved. After mitosis, Cdc20 is degraded and the APC/C interacts with the Cdc20-related protein Cdh1. Cdh1 interacts with the site occupied by the autoinhibitory segment of Apc1, and yet phosphorylation of this segment is not required for Cdh1 binding. How does Cdh1 bind this region? Cdh1 might be a better competitor than Cdc20 for binding at this site[8], or there might be other mechanisms at play.

Finally, the Apc1 loop and autoinhibitory segment are poorly evolutionarily conserved outside vertebrates, raising questions about APC/C regulation in other species. Fortunately, addressing these and related problems has just become much easier, now that we have a strong structural foundation on which to base future experiments. We are a major step closer to understanding the remarkable robustness of mitosis. ■

David O. Morgan *is in the Department of Physiology, University of California, San Francisco, California 94158, USA.*
e-mail: david.morgan@ucsf.edu

1. Musacchio, A. *Curr. Biol.* **25,** R1002–R1018 (2015).
2. Hoyt, M. A., Totis, L. & Roberts, B. T. *Cell* **66,** 507–517 (1991).
3. Li, R. & Murray, A. W. *Cell* **66,** 519–531 (1991).
4. Alfieri, C. *et al. Nature* **536,** 431–436 (2016).
5. Yamaguchi, M. *et al. Mol. Cell* http://dx.doi.org/10.1016/jmol.cel.2016.07.003 (2016).
6. Primorac, I. & Musacchio, A. *J. Cell Biol.* **201,** 177–189 (2013).
7. Izawa, D. & Pines, J. *Nature* **517,** 631–634 (2015).
8. Zhang, S. *et al. Nature* **533,** 260–264 (2016).
9. Yang, Q. & Ferrell, J. E. Jr *Nature Cell Biol.* **15,** 519–525 (2013).
10. Fujimitsu, K., Grimaldi, M. & Yamano, H. *Science* **352,** 1121–1124 (2016).
11. Qiao, R. *et al. Proc. Natl Acad. Sci. USA* **113,** E2570–E2578 (2016).
12. Valk, E. *et al. Mol. Biol. Cell* **25,** 3456–3460 (2014).

This article was published online on 10 August 2016.

ASTRONOMY

Earth-like planet around Sun's neighbour

An Earth-mass planet has been discovered in orbit around Proxima Centauri, the closest star to our Sun. The planet orbits at a distance from the star such that liquid water and potentially life could exist on its surface. SEE LETTER P.437

ARTIE P. HATZES

Exoplanet discoveries trigger our imaginations. What does the newly discovered planet look like and what are its characteristics? Could it harbour life and, if so, how many such habitable worlds are there in our Galaxy? Astronomers aim to find Earth-like planets in the temperate zone of a star — the distance from the star at which the planet's surface temperature could theoretically support liquid water. On page 437, Anglada-Escudé *et al.*[1] report the exciting discovery of an Earth-mass planet in the temperate zone of Proxima Centauri, the closest star to our Sun.

Proxima Centauri is a low-mass star (it has 12% of the Sun's mass[1]) that belongs to the family of stars known as M dwarfs. Astronomers are focusing their efforts on finding small, potentially habitable planets around M dwarfs because such planets can be discovered with the instruments available today. An Earth-like planet orbiting in the temperate zone can be detected through a 'Doppler wobble' — the effect caused by the planet's gravitational tug on the motion of its host star. This method was used to discover the first exoplanet around a Sun-like star[2] in 1995.

Applying the Doppler method to an M dwarf has two benefits. First, such stars are cooler than the Sun, which means that their temperate-zone planets would orbit much closer to the star (about one-tenth of the Sun–Earth distance[3]). Second, these stars are less massive than the Sun. Both of these factors result in a Doppler wobble for M dwarfs that is large enough (about 1 metre per second) to be detected by modern instruments[4]. By comparison, Earth induces a Doppler wobble in the Sun of 0.09 m s^{-1}.

M dwarfs are the most abundant stars in the Galaxy, but so far, only a few Earth-mass planets have been discovered in the temperate zones of such stars[5]. If only a small fraction of M dwarfs have temperate-zone planets, our Galaxy could be teeming with life.

Anglada-Escudé and colleagues first found their exoplanet, which is called Proxima Centauri b, using Doppler measurements taken by the Ultraviolet and Visual Echelle Spectrograph at the European Southern Observatory (ESO) in Chile between 2000 and 2008. These data showed a hint — but not an entirely convincing one — of a Doppler wobble of 1.38 m s^{-1}. The authors confirmed the signal by using many more Doppler measurements taken in 2016 with the ESO's High Accuracy Radial velocity Planet Searcher.

The authors' careful analysis of the data eliminated other possible causes for the Doppler wobble, such as stellar activity. The signal is particularly convincing because it is seen in two independent data sets that span more than a decade. In addition to the fact that the planet is orbiting Proxima Centauri, there are two further exciting aspects of Anglada-Escudé and colleagues' discovery: the orbital period of 11.2 days places the planet within

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

ESO/M. KORNMESSER



Figure 1 | Artist's impression of the exoplanet Proxima Centauri b. Anglada-Escudé *et al.*[1] have discovered an Earth-like planet in orbit around the closest star to our Sun, Proxima Centauri. The planet's surface temperature should allow it to support liquid water, and its mass suggests that it might have a rocky surface. In this artist's impression, the planet is assumed to have an Earth-like atmosphere.

the temperate zone of Proxima Centauri, and the planet is only 1.3 times more massive than Earth, so it might be Earth-like (Fig. 1). These observations naturally raise the question of whether Proxima Centauri b could harbour life. However, circling a star at the right distance is no guarantee that the planet has liquid water, or even an atmosphere that can support life.

One problem for Proxima Centauri, and for M dwarfs in general, is that these stars are much more active than the Sun. Proxima Centauri produces powerful flares[6], and the X-ray flux received by the planet is 400 times greater than the flux that Earth receives from the Sun. Energetic particles associated with the flares may erode the atmosphere or hinder the development of primitive forms of life. We also don't know whether the exoplanet has a magnetic field, like Earth, which could shield it from the dangerous stellar radiation. Until we understand what makes a planet habitable, it is better to say that Proxima Centauri b lies in a temperate zone (the right temperature) rather than a habitable zone (the right conditions to support life). Interestingly, M-dwarf stars are long-lived, and Proxima Centauri will exist for several hundreds or thousands of times longer than the Sun[7]. Any life on the planet could still be evolving long after our Sun has died.

Studies of the exoplanet's atmosphere could assess its habitability. One way to do this is by using a technique called transmission spectroscopy[8]: if the planet passes in front of (transits) the star when viewed from Earth, its atmosphere would absorb the starlight while transiting, and so the spectrum of the starlight would contain a signature of the planet's atmosphere. But we currently do not know whether Proxima Centauri b is such a transiting planet — there is only a 1.5% chance that it is[1].

One could also attempt to detect the reflected or radiated light from the planet directly[9] — this can be done only for nearby planets. Because Proxima Centauri is relatively close to us, such attempts have a reasonable chance of succeeding. In the distant future, an interstellar space probe might get a close-up look at the planet.

NASA's Transiting Exoplanet Survey Satellite, which is scheduled for launch in 2017, will search for transiting planets around thousands of the closest M dwarfs[10]. The planets' atmospheres could then be characterized using the James Webb Space Telescope, scheduled for launch in 2018. In the next decade, we will learn much more about the atmospheres of exoplanets in the temperate zones of M-dwarf stars. Meanwhile, Proxima Centauri b gives astronomers their best opportunity yet to study such planets. ■

Artie P. Hatzes *is at the Thuringian State Observatory, Tautenburg 07778, Germany.*
e-mail: artie@tls-tautenburg.de

1. Anglada-Escudé, G. *et al. Nature* **536,** 437–440 (2016).
2. Mayor, M. & Queloz, D. *Nature* **378,** 355–359 (1995).
3. Kasting, J. F., Whitmire, D. P. & Reynolds, R. T. *Icarus* **101,** 108–128 (1993).
4. Pepe, F. *et al. Astron. Astrophys.* **534,** A58 (2011).
5. Gillon, M. *et al. Nature* **533,** 221–224 (2016).
6. Güdel, M., Audard, M., Reale, F., Skinner, S. L. & Linsky, J. L. *Astron. Astrophys.* **416,** 713–732 (2004).
7. Adams, F. C. & Lauglin, G. *Rev. Mod. Phys.* **69,** 337–372 (1997).
8. Bean, J. L., Kempton, E. M.-R. & Homeier, D. *Nature* **468,** 669–672 (2010).
9. de Kok, R. J. *et al. Astron. Astrophys.* **554,** A82 (2013).
10. Ricker, G. R. *et al. J. Astron. Telesc. Instrum. Syst.* **1,** 014003 (2015).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

ARTICLE

doi:10.1038/nature19082

Early onset of industrial-era warming across the oceans and continents

Nerilie J. Abram[1,2], Helen V. McGregor[3], Jessica E. Tierney[4,5], Michael N. Evans[6], Nicholas P. McKay[7], Darrell S. Kaufman[7] & the PAGES 2k Consortium*

The evolution of industrial-era warming across the continents and oceans provides a context for future climate change and is important for determining climate sensitivity and the processes that control regional warming. Here we use post-AD 1500 palaeoclimate records to show that sustained industrial-era warming of the tropical oceans first developed during the mid-nineteenth century and was nearly synchronous with Northern Hemisphere continental warming. The early onset of sustained, significant warming in palaeoclimate records and model simulations suggests that greenhouse forcing of industrial-era warming commenced as early as the mid-nineteenth century and included an enhanced equatorial ocean response mechanism. The development of Southern Hemisphere warming is delayed in reconstructions, but this apparent delay is not reproduced in climate simulations. Our findings imply that instrumental records are too short to comprehensively assess anthropogenic climate change and that, in some regions, about 180 years of industrial-era warming has already caused surface temperatures to emerge above pre-industrial values, even when taking natural variability into account.

Palaeoclimate data from the past two millennia—a period for which natural and anthropogenic climate forcings are reasonably well constrained—provide perspectives on global temperature changes during the twentieth century. Climate reconstructions of the past 2,000 years have focused mainly on the Northern Hemisphere[1–3], using records derived primarily from terrestrial settings. Recent continental-scale temperature reconstructions provide evidence for twentieth century warming over all reconstruction regions except Antarctica[4]. A new Southern Hemisphere temperature reconstruction also demonstrates that the twentieth century is the only period of the last millennium during which warm extremes occurred simultaneously across both hemispheres[5]. However, these hemispheric and regional temperature histories do not allow for assessments of how past temperature changes evolved between the oceans and land.

The oceans represent a major heat reservoir, taking up more than 90% of the total global energy imbalance since the 1950s[6]. Internal variability of ocean circulation mediates the global climate and, for example, is implicated in the slowdown of global atmospheric warming during the 'hiatus' interval[7] (AD 2001–2014) because of the drawdown of additional heat into the subsurface ocean[7,8]. The existence of the recent warming slowdown is debated[9]; however, earlier decade-scale plateaus in the rate of warming are prominent features of the climate record[10]. Given the importance of the oceans in determining the pace and regional structure of changes in climate[11], it is essential to understand how anthropogenic warming developed in the oceans and over land during the industrial era.

Determining an unambiguous time for the start of the industrial era is difficult, and forms part of the debate over a formal definition of the Anthropocene[12,13]. The Intergovernmental Panel on Climate Change (IPCC) uses 'industrial era' to refer, somewhat arbitrarily, to the time after AD 1750, when industrial growth began in Britain, spread to other countries and led to a strong increase in fossil fuel use and greenhouse gas emissions. Here, we use the term 'industrial-era warming' to refer to the sustained, significant ($P < 0.1$) warming of Earth's climate that developed during the industrial era. We use the palaeoclimate history since AD 1500 as a context for assessing the evolution of industrial-era warming across surface-ocean and land areas. Our assessments use newly developed regional sea-surface temperature (SST) reconstructions for the tropical oceans[14] and SST-sensitive records for the global oceans[14,15], along with continental-scale temperature reconstructions and databases[4,16] (Fig. 1, Methods). We compare the onset of industrial-era warming in these palaeoclimate datasets to transient multi-model climate simulations driven by full natural and anthropogenic forcings[17]. We also use experiments with single[18–20] and cumulative[21] external climate forcings to investigate the factors that define the onset of industrial-era warming.

Regional features of industrial-era warming

Synthesis of marine palaeoclimate records spanning the past 2,000 years has identified a robust global surface-ocean cooling trend that reached coolest conditions during the period AD 1400–1800[15]. This finding is qualitatively consistent with pre-industrial cooling trends in terrestrial records and can be explained by an increased frequency of explosive volcanism during the past millennium[15,22]. Marine records with moderate-to-high (<25 yr) temporal resolution indicate that, in many regions, this long-term SST cooling trend reversed during the industrial era[14,15], including in the tropical oceans where robust regional SST reconstructions spanning the past four centuries have been developed using coral archives[14]. Industrial-era warming in the area-weighted average of regional tropical SST reconstructions is visually similar to warming of the global area-weighted mean of terrestrial temperature reconstructions[4] (Fig. 1c, d). In particular, the average

[1]Research School of Earth Sciences, Australian National University, Canberra, Australian Capital Territory 2601, Australia. [2]ARC Centre of Excellence for Climate System Science, Australian National University, Canberra, Australian Capital Territory 2601, Australia. [3]School of Earth and Environmental Sciences, University of Wollongong, Wollongong, New South Wales 2522, Australia. [4]University of Arizona, Department of Geosciences, Tucson, Arizona 85721, USA. [5]Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA. [6]Department of Geology and Earth System Science Interdisciplinary Center, University of Maryland, College Park, Maryland 20742, USA. [7]School of Earth Sciences and Environmental Sustainability, Northern Arizona University, Flagstaff, Arizona 86011, USA.
*A list of participants and their affiliations appears at the end of the paper.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

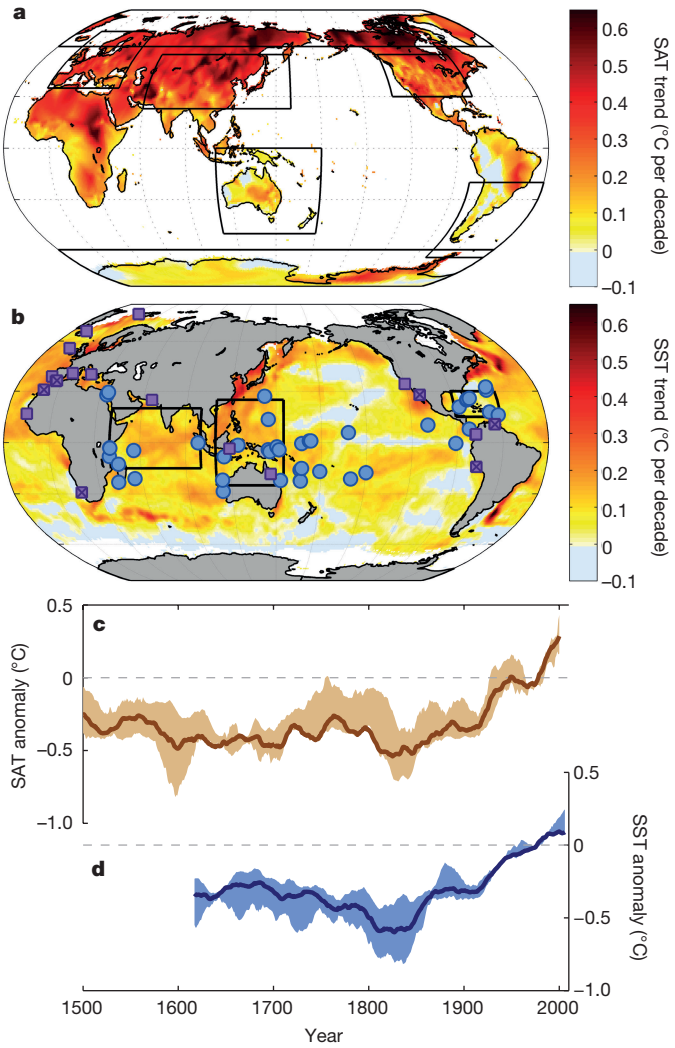


Figure 1 | Terrestrial and marine palaeoclimate reconstructions. a, **b**, Instrumental temperature trends over the period AD 1961–2010 for surface air temperature (SAT; **a**) and sea surface temperature (SST; **b**). The black boxes indicate the PAGES 2k continental-scale reconstruction regions[4] (**a**) and the PAGES Ocean2k high-resolution (annual) reconstruction regions[14] (**b**). The symbols in **b** indicate the locations of highly resolved[14] (blue circles; Extended Data Table 1) and moderately resolved[15] (<25 yr, purple squares; crossed squares indicate ocean upwelling sites; Extended Data Fig. 7) marine palaeo-SST proxy records. **c**, **d**, 25-year moving averages of the area-weighted mean terrestrial temperature anomaly (brown line)[4,16] and the 25%–75% range across continental-scale reconstructions (shading) (**c**) and area-weighted mean tropical SST anomaly (blue line) and minimum–maximum range across the Indian, western Pacific and western Atlantic reconstructions[14] (shading) (**d**). Anomalies are relative to the AD 1961–1990 mean (dashed lines).

terrestrial and tropical ocean temperature histories show industrial-era warming developing after AD 1800, with similar multi-decadal expressions of accelerated and reduced warming phases.

The similarity of average terrestrial and marine temperature histories (Fig. 1c, d) masks important regional differences in industrial-era warming (Fig. 2, Extended Data Fig. 1). To examine these regional features, we first assess when sustained, significant warming began in the regional temperature reconstructions. We define a sustained, significant trend as the most recent trend that persists until the end of the reconstruction and that is significantly different from zero above the 90% confidence level ($P < 0.1$). We determine the median time of onset of these sustained trends across different levels of smoothing (15–50 y) applied to the regional reconstructions (Methods, Extended Data Fig. 2a)[23,24]. The strengths and limitations of this and other change-point detection methods in assessing the onset of industrial-era warming are explored using synthetic time series (Methods, Extended Data Fig. 3). Model-based testing suggests that estimates for warming onset are insensitive to the seasonal preference that exists in some regional reconstructions[4] (Methods, Extended Data Fig. 4).

Sustained, significant warming began in the tropical oceans around the 1830s, with no discernible difference in onset across the three tropical SST reconstruction regions (Fig. 2a, Table 1, Extended Data Fig. 1a, b). The onset of tropical-ocean warming is similar to the median onset of warming in Northern Hemisphere mean temperature[3] (Extended Data Fig. 1a), although the Northern Hemisphere ensemble also includes a sub-member where recent warming is not sustained to the end of the reconstruction (Supplementary Fig. 1). This exception is due to strong multi-decadal variability, which is common to each of the Northern Hemisphere mid-latitude continental regions and can delay the detection of sustained warming trends at narrow filter widths (Extended Data Fig. 5). Nevertheless, each of the Northern Hemisphere regional-scale reconstructions also displays mid-nineteenth-century onsets for industrial-era warming (Fig. 2a, Table 1, Extended Data Fig. 1a, b).

In contrast to the mid-nineteenth-century onset of warming in the tropical oceans and the Northern Hemisphere, the Southern Hemisphere onset of industrial-era warming appears delayed (Fig. 2a, Table 1). In hemispheric-scale reconstructions[3,5], the median estimate for the onset of sustained warming is approximately 50 years later in the Southern Hemisphere than the Northern Hemisphere (Extended Data Fig. 1a, Supplementary Fig. 1). The regional structure of this apparent Southern Hemisphere lag involves sustained, significant warming developing over Australasia and South America around the start of the twentieth century, while significant continent-scale warming is not detected for Antarctica (Table 1, Extended Data Fig. 1a). The Antarctic reconstruction has the greatest uncertainty of the regional reconstructions[4] (Methods), and significant warming has been documented over the Antarctic Peninsula and West Antarctica since the mid-twentieth century[23,25]. However, the absence of significant Antarctic warming at the continent scale is corroborated by post-1979 satellite observations averaged across Antarctica[26].

Industrial-era warming across the oceans and continents is further investigated by examining the rates of regional warming. All century-scale linear trends in the regional reconstructions since AD 1500 are calculated (100-yr trends with 1-yr time step; Supplementary Video 1) and the distributions of trends beginning since AD 1800 are used to assess the regional rates of industrial-era warming. For each tropical ocean and Northern Hemisphere regional reconstruction, the distribution of century-scale trends starting after AD 1800 has a clear positive shift compared to earlier trends (Fig. 2b, Table 1, Extended Data Fig. 1c) and includes the largest century-scale warming trend of the past 500 years. Temperature trends in the Arctic since AD 1800 are greater than in any other region, indicative of Arctic amplification[27]. The similarity of post-AD 1800 trends for the tropical Indian and western Atlantic oceans with those in Europe, Asia and North America (Table 1), indicates that industrial-era warming of the tropical oceans has progressed at a rate similar to warming of the Northern Hemisphere mid-latitude continents. By contrast, rates of century-scale warming since AD 1800 in the Southern Hemisphere regional reconstructions are slower than for the tropical oceans and Northern Hemisphere continents (Fig. 2b, Table 1). This difference may be related to the delayed onset of warming in the Australasia and South America reconstructions, but is also consistent with instrumental evidence for hemispheric asymmetries in the rate of twentieth-century warming (Fig. 1a, b). For Antarctica, the absence of continent-scale warming during the industrial era results in pre- and post-AD 1800 trend distributions that are statistically indistinguishable (Extended Data Fig. 1c).

The time when a climate-change signal exceeds the range of climate variability is known as the 'time of emergence'[28]. The time of emergence

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

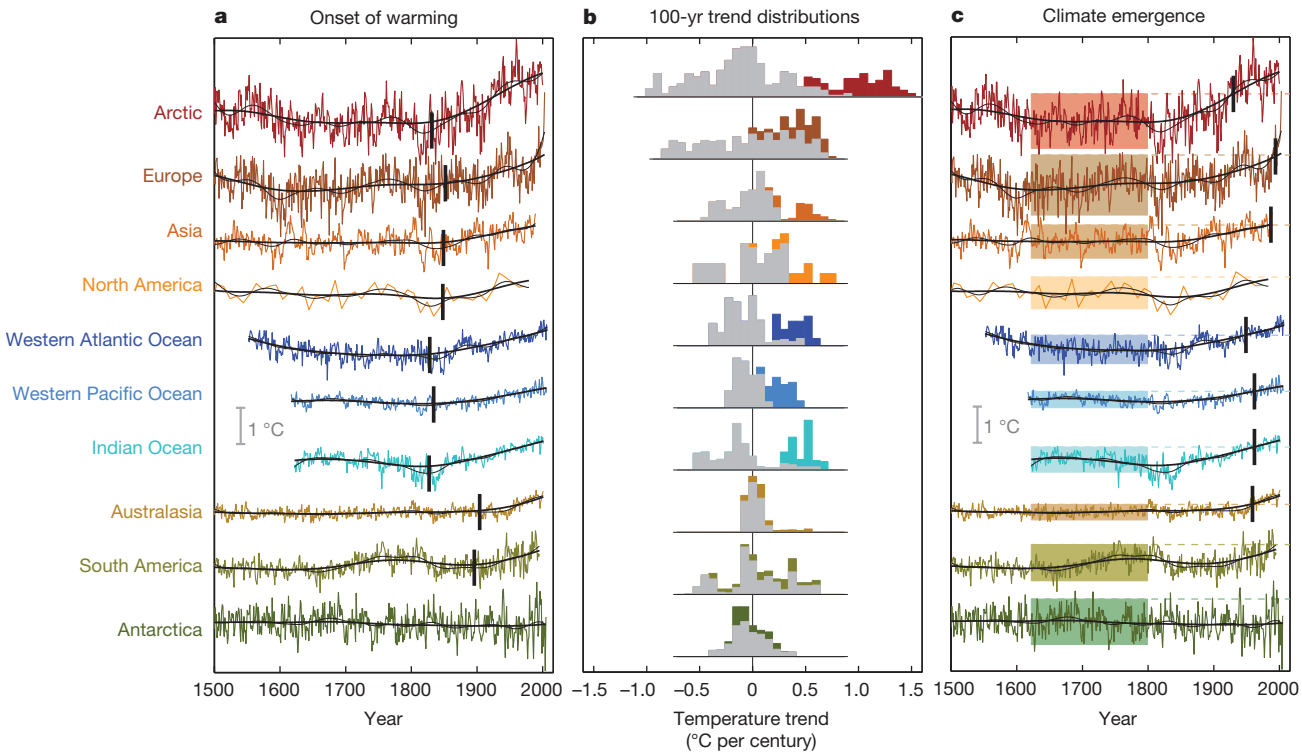


Figure 2 | Onset and magnitude of industrial-era warming in regional temperature reconstructions. a, Regional reconstructions[4,14] since AD 1500 (coloured lines) with 15-yr (thin black lines) and 50-yr (thick black lines) Gaussian smoothing, shown alongside the median time of onset for sustained, significant industrial-era warming assessed across 15–50-yr filter widths (vertical black bars) (Methods, Extended Data Fig. 2a). Grey 1 °C scale bar denotes the *y*-axis scale of each regional temperature reconstruction. **b**, Histograms of century-scale regional trends, comparing those beginning during the period AD 1500–1799 (grey bars) with those beginning since AD 1800 (coloured bars). The heights of the bars are normalized by the maximum occurrence across each region to aid visual representation. See Supplementary Video 1 for temporal evolution of century-scale trends. **c**, Regional reconstructions (coloured lines) with 15-yr (thin black lines) and 50-yr (thick black lines) filters, shown alongside the $\pm 2\sigma$ range (shaded boxes) of interannual variability over the AD 1622–1799 reference period ($+2\sigma$ level continued by dashed lines). The median time of climate emergence (vertical black bars) is defined as the time at which the climate signal (15–50-yr width filters of the regional reconstructions) first exceeds and remains above the $+2\sigma$ threshold of the reference period (Methods).

for industrial-era warming depends on (i) when warming began, (ii) the rate of warming and (iii) the magnitude of interannual to multi-decadal climate variability. Time-of-emergence studies typically use twentieth century instrumental data or post-1850 (historical) simulations to characterize the baseline climate, and have commonly concluded that unprecedented climates will emerge first in tropical air temperatures because of the small magnitude of interannual temperature variability in these regions[28–30]. However, our findings of a mid-nineteenth-century onset of industrial-era warming suggest that, in some regions, the entire instrumental period contains a signature of climate warming, rendering it unsuitable for determining climate emergence. We use the multi-century context available from the regional palaeoclimate reconstructions, and a pre-AD 1800 reference period (Methods), to assess the extent to which industrial-era warming may have already emerged in regional climates.

Industrial-era warming led to the emergence of regional climate change first in the Arctic (Fig. 2c, Table 1). Despite the large variability of Arctic climate, the palaeoclimate time-of-emergence assessment indicates that the early onset and rapid rate of warming resulted in the emergence of climate change during the 1930s (approximately 100 years after sustained, significant warming began). The tropical ocean regions display a similar rate of warming to that of the Northern Hemisphere mid-latitude continents, but the industrial-era warming signal emerges sooner in the tropical oceans (time of emergence at around AD 1948–1962) because of the smaller magnitude of variability there. Emergence of industrial-era warming for Australasia is around AD 1960, because the delayed onset of warming is compensated by the small magnitude of interannual variability in this regional reconstruction. All other regions apart from Antarctica are nearing the emergence of warming above the threshold of pre-industrial climate variability by the start of the twenty-first century (the end of the reconstructions).

Our regional palaeoclimate assessments suggest that widespread climate warming observed during the twentieth century forms part of a sustained trend that began in the tropical oceans and over some Northern Hemisphere land areas around the 1830s (about 180 years ago). Although caveats exist relating to the accuracy with which regional palaeoclimate reconstructions represent past temperature changes[4,14–16], our multi-century assessments clearly demonstrate the need to incorporate pre-twentieth-century information in comprehensive assessments of industrial-era warming. The early onset of industrial-era warming may not alter the conclusions of time-of-emergence studies focused on protecting infrastructure built during recent decades[31]; however, our findings imply that time-of-emergence studies that rely on a twentieth-century baseline may underestimate how soon the effects of climate change will fall outside the range of climate variability to which natural systems are adapted[30].

Climate forcing of industrial-era warming

Model simulations provide an important tool for investigating which forcings are most consistent with the reconstructed onset of industrial-era warming. We examine the regional responses of global climate model simulations to natural and anthropogenic forcings since AD 1500, applying the same trend-detection methodology used for the palaeoclimate reconstructions (Methods). An ensemble of ten different models reproduces the near-synchronous mid-nineteenth-century onset of sustained, significant warming observed for reconstructed Northern Hemisphere surface air temperature and tropical SST

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

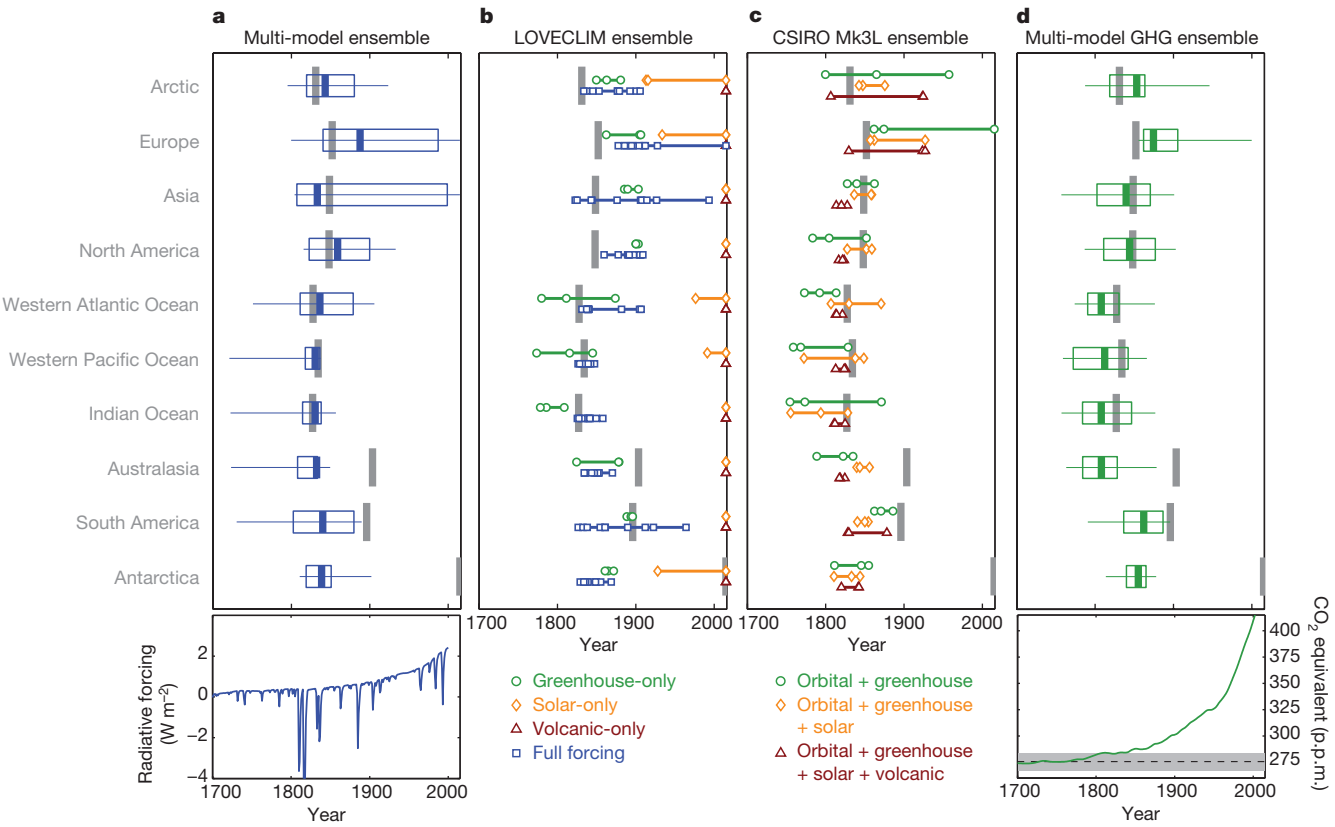


Figure 3 | Data–model comparison of the onset of industrial-era warming. a–d, Median onset of sustained, significant warming in regional reconstructions (vertical grey bars; as in Fig. 2a, vertical black bars) compared to various ensemble results. **a**, Median (vertical blue bars), 25%–75% range (boxes) and 5%–95% range (horizontal lines) of the corresponding regional median warming onsets across ten multi-model last-millennium climate simulations with full radiative forcings (Supplementary Fig. 1). **b**, Median timing (open symbols) of warming onset across three-member ensembles for single radiative forcing experiments (green circles, orange diamonds and red triangles) and a ten-member ensemble with full radiative forcings (blue squares) using the LOVECLIM model[18]. **c**, Median timing (open symbols) of warming onset for three-member ensembles of CSIRO Mk3L simulations using cumulative radiative forcing experiments[21]. **d**, Median (green vertical bars), 25%–75% range (boxes) and 5%–95% range (horizontal lines) of regional median warming onsets for 13 ensemble members across four different climate models forced with only greenhouse gas changes[18–21]. Lower panel in **a** shows the corresponding full radiative forcing (from orbital, greenhouse gas, solar and volcanic sources)[17]. Lower panel in **d** shows the corresponding equivalent atmospheric CO_2 concentration relative to the mean (dashed line) and ±1.5 interquartile range (grey shading; outlier test) of an AD 0–1500 baseline interval. See Extended Data Table 2 for model details.

(Fig. 3a, Table 1). Palaeoclimate data–model agreement is particularly good for Northern Hemisphere terrestrial regions, where the patterns of short-term cooling caused by volcanic eruptions and sustained recent warming from greenhouse gas emissions are remarkably similar (Extended Data Fig. 2). The agreement between the multi-model ensemble and palaeoclimate reconstructions suggests that the onset of industrial-era warming over Northern Hemisphere landmasses and in the tropical oceans is consistent with a forced climate response.

By contrast, none of the climate models show evidence for a delayed Southern Hemisphere onset of industrial-era warming (Fig. 3a). A ten-member ensemble of LOVECLIM simulations also suggests that the delayed development of warming over Antarctica and Australasia is not explicable within the range of unforced climate variability in that model (Fig. 3b). Instead, the evolution of regional temperature trends in the multi-model ensemble mean (Extended Data Fig. 2b) and within individual models (Supplementary Fig. 2) shows a globally synchronous thermodynamic response of surface temperatures to external climate forcings. Previous palaeoclimate studies have noted that climate models tend to overemphasize Northern Hemisphere–Southern Hemisphere synchronicity of past temperature changes[5,32], perhaps by overestimating externally forced climate responses in the Southern Hemisphere and/or underestimating the magnitude of Southern Hemisphere climate variability[32]. Unresolved or inaccurate physical processes in model representations of the Southern Hemisphere ocean–atmosphere–cryosphere systems also hinder the accurate simulation of Southern Hemisphere climate[26,33] including the overestimation of simulated Antarctic region warming compared with satellite observations of recent surface air and ocean temperature trends[26]. The paucity of climate observations also hinders attempts to resolve differences between observations and simulations in the Southern Hemisphere[26,32]. As a result, currently available model output cannot reasonably be used to assess the delayed onset of Southern Hemisphere industrial-era warming that is suggested by palaeoclimate observations.

Naturally forced climate cooling may have helped to set the stage for the widespread onset of industrial-era warming in the tropical oceans and over Northern Hemisphere landmasses during the mid-nineteenth century. Episodic cooling caused by the large 1815 Tambora volcanic eruption is prominent in Northern Hemisphere terrestrial temperature reconstructions (Fig. 2a). In last-millennium model simulations, the strong cooling caused by the Tambora eruption is followed immediately by a decade-scale interval of accelerated global warming as the climate recovers[34]. The Dalton solar minimum also occurred in the early nineteenth century, but solar forcing is thought to have only a small influence on last-millennium climate compared to the effects of volcanic eruptions[20]. In the CSIRO Mk3L experiments, in which forcings were applied cumulatively rather than individually, it is particularly the addition of volcanic forcing that tends to focus the onset of industrial-era warming into a narrower time window—both between regions and

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Table 1 | Onset, rate and emergence of industrial-era warming in reconstructions and simulations

	Year of onset of sustained, significant warming trends		Century-scale trend distribution (°C per century)		Year of emergence
	Reconstructions	Simulations	Reconstructions		Reconstructions
			AD 1500–1799	Since AD 1800	
Arctic	1831	1843 (1819–1880)	−0.11 (−0.41–0.07)	1.07 (0.92–1.26)	1930
Europe	1852	1888 (1840–1987)	0.02 (−0.39–0.33)	0.46 (0.29–0.58)	1994‡
Asia	1849	1833 (1807–1999)	0.00 (−0.18–0.10)	0.48 (0.35–0.54)	1987‡
North America	1847	1859 (1823–1900)	0.01 (−0.38–0.15)	0.48 (0.29–0.52)	§
Western Atlantic Ocean	1828	1836 (1811–1879)	−0.11 (−0.21–0.01)	0.41 (0.27–0.51)	1948
Western Pacific Ocean	1834	1830 (1818–1836)	−0.09 (−0.13–0.02)	0.27 (0.19–0.35)	1962
Indian Ocean	1827	1830 (1814–1838)	−0.14 (−0.31–−0.04)	0.51 (0.39–0.54)	1962
Australasia	1904*	1832 (1808–1833)	0.00 (−0.04–0.05)	0.07 (0.02–0.23)	1959
South America	1896	1840 (1802–1880)	−0.02 (−0.11–0.21)	0.20 (−0.02–0.38)	§
Antarctica	†	1839 (1819–1851)	−0.05 (−0.15–0.09)	−0.06 (−0.14–0.07)	§

Statistics represent changes in surface air temperature (SAT) for continental reconstruction regions and sea surface temperature (SST) for tropical ocean regions. The values given are medians (inter-model medians for the simulations), with the 25%–75% ranges in parentheses. The reconstructed median years of onset and emergence are determined using 15–50-yr filter widths. Visual representations of the data are found in Fig. 2a (onset of warming in reconstructions), Fig. 3a (onset of warming in simulations), Fig. 2b (rates of warming) and Fig. 2c (climate emergence).
*Compared with a median warming onset of AD 1886 in the original Australasia 2k reconstruction[4] that includes marine SST-sensitive records.
†Sustained, significant warming never achieved in reconstruction.
‡Emergence of industrial-era warming above the AD 1622–1799 (reference interval) variability was within 20 years of the end of the reconstruction, making permanent emergence uncertain.
§Emergence was not achieved in reconstruction.

between ensemble members for the same region—compared to experiments run with greenhouse forcing alone (Fig. 3c).

Simulations suggest that recovery from volcanic cooling is not an essential requirement for reproducing the mid-nineteenth-century onset of industrial-era warming. Multi-model experiments forced with only greenhouse gases capture regional onsets for sustained industrial-era warming that are consistent with the tropical ocean and Northern Hemisphere continental reconstructions (Fig. 3d). Testing of our change-point detection method also indicates that volcanic-style cooling events do not substantially alter the onset determined for sustained warming trends in synthetic time series (Methods, Extended Data Fig. 3a, b). We conclude that greenhouse forcing of industrial-era warming began by the mid-nineteenth century, but that the confluence of explosive volcanic eruptions around this time was probably also influential in aligning the onset of warming over tropical ocean and Northern Hemisphere land regions.

Mechanisms of industrial-era warming

Our regional palaeoclimate assessments show that the thermodynamic response to increasing greenhouse gas concentrations developed in the oceans and atmosphere even when anthropogenic contributions were small. The spatial fingerprint of the onset of industrial-era warming may further elucidate the role of the oceans in the development of anthropogenic warming. We explore this by assessing change-points in the site-level palaeoclimate records that contribute to the marine[14,15] and terrestrial[4,16] temperature reconstructions (Extended Data Fig. 6). Onset estimates from site-level records are more variable than for the regional reconstructions. This variation arises from (i) lower trend-to-variability (or signal-to-noise) ratios in individual palaeoclimate records, (ii) varying lengths of the individual records that do not always include information before the onset of sustained industrial-era trends and (iii) differences in how representative each record is of local temperature. Issues similar to (i) and (ii) limit the detection and attribution of climate change at subregional scales from climate observations and simulations[35]. For these reasons, we view the site-level analyses in only a qualitative sense (Fig. 4), using them to aid the more robust assessments derived from regional reconstructions (Fig. 2).

Sustained, significant warming trends developed during the industrial era across the majority (71%, $n = 55$) of marine records (Extended Data Fig. 6a). This complements similar findings of widespread recent warming trends in site-level palaeoclimate records from predominantly terrestrial environments[4]. Development of sustained warming during the nineteenth century in individual records from the tropical oceans and in Northern Hemisphere mid- and high-latitude terrestrial records (Fig. 4) corroborate our findings based on the regional reconstructions. Some Southern Hemisphere mid-latitude terrestrial records also show this early warming (Fig. 4a), suggesting that although continental-scale temperature reconstructions for Australasia and South America indicate a delayed onset of industrial-era warming (Fig. 2), this may not be representative of all Southern Hemisphere mid-latitude land areas (Fig. 4a).

The early onset of industrial-era warming of the tropical oceans was widespread; however, industrial-era climate changes may have resulted in localized surface-ocean cooling in some settings (Extended Data Figs 6b, 7a). Previous assessments of the moderately resolved marine records cautiously concluded that qualitative warming and cooling trends during the twentieth century were produced if the records were composited into *a priori* defined non-upwelling and upwelling subsets, or if the records were composited by tropical versus Northern Hemisphere extratropical location, or by proxy type[15]. Our change-point assessment of the moderately resolved marine records finds that the most distinct differentiation between recent cooling and warming trends occurs when the records are separated into upwelling and non-upwelling sites (Extended Data Fig. 7b–d). Therefore, enhanced ocean upwelling may be the most plausible mechanism for explaining the recent cooling trends detected at some marine sites, consistent with theories that climate warming could, in some locations, cause strengthening of the surface winds that generate coastal upwelling[36–38].

The early development and rapid rate of tropical ocean warming during the industrial era (Figs 2, 4b) may corroborate model-based descriptions of an enhanced equatorial response of the oceans to increased greenhouse forcing[39,40]. The warming near the equator as a result of an enhanced equatorial response is caused by increased surface ocean stability and a reduction in surface evaporative cooling due to the combination of lower wind speed and relative humidity. The hypothesized enhanced-equatorial-response mechanism and spatial fingerprint differs from an El Niño-like response of the tropical oceans to global warming (requiring weaker Walker circulation, reduced eastern Pacific upwelling and reduced east–west SST gradient)[41] and shows greater consistency between models[39]. A regional SST reconstruction for the eastern Pacific, although not used in our study because it is thought to have a spuriously large twentieth-century trend attributed to hydrologic effects[14], indicates that sustained, significant warming of eastern Pacific SST began markedly later (around AD 1913) than in the other tropical ocean regions[14]. Full-forcing and greenhouse-only climate model simulations assessed spatially at a grid resolution of 5° × 5° (Extended Data Fig. 8) also display a delayed onset of industrial-era warming in the eastern Pacific over a narrow region along the

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

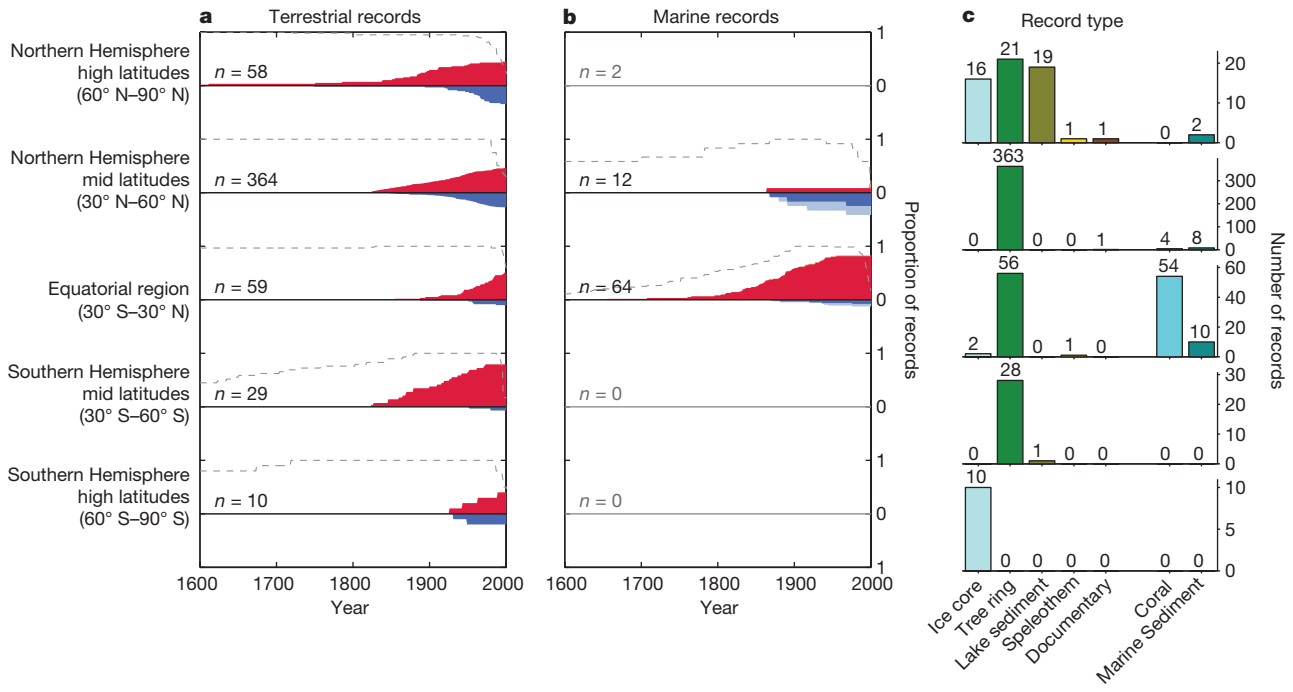


Figure 4 | Latitudinal development of site-level temperature trends. **a**, **b**, Cumulative distributions for terrestrial (**a**) and marine (**b**) records, showing latitudinal development of sustained significant warming (red; upward) or cooling (blue; downward) trends across the site-level records (Extended Data Fig. 6). The distributions are expressed as a proportion of total records within each latitudinal band (number of total records, *n*, indicated; grey values indicate an insufficient number of site-level records ($n \leq 2$) for meaningful comparison). Light shading in **b** denotes trends at marine sediment core sites with an *a priori* upwelling regime (Extended Data Fig. 7), for which the localized SST response[36–38] to regional climate warming may not be representative of latitudinal average climate. Dashed lines show the temporal coverage of site-level records (expressed as a proportion of latitudinal band total, *n*). **c**, The number and type of proxy records available by latitudinal band (bars are coloured by archive type). Because of the varying length of individual records, not all of which cover the full interval of industrial-era warming (particularly tropical corals), we use these data only as a qualitative indicator of the relative timing of the onset of industrial-era warming between different latitudinal bands.

equator, distinct from the otherwise widespread early onset of tropical surface ocean warming (Extended Data Fig. 8, Supplementary Fig. 3). The delayed detection of sustained, significant warming in the eastern Pacific may be influenced by large-amplitude interannual variability there (Extended Data Fig. 3b, c), which limits our ability to confidently evaluate El Niño (or La Niña)-like dynamic mechanisms in defining the onset of industrial-era warming across the oceans. Nevertheless, the early onset of industrial-era warming in tropical ocean regions away from the equatorial eastern Pacific, in reconstructions and simulations, appears to support an enhanced equatorial response in the SST changes caused by global warming.

Widespread warming of tropical SSTs during the industrial era (Figs 2, 4b) may have had global importance because of the non-linear influence of tropical SSTs on deep atmospheric circulation, which redistributes heat and moisture. The latitudinal development of terrestrial warming that we observe across individual palaeoclimate records (Fig. 4a) is similar to that reported for twentieth-century terrestrial air temperature observations[42]. In that study it was proposed that twentieth-century terrestrial warming focused over the Northern Hemisphere subtropical–subpolar regions and in a narrow band over the Southern Hemisphere subtropics could be indicative of anthropogenically forced widening of the tropics through expansion of the Hadley circulation cells[42]. The causes of recent multi-decadal episodes of contraction and expansion of Hadley circulation[43,44] have proven difficult to resolve because of the brevity of observational datasets. However, numerous studies suggest that tropical ocean warming is essential for reproducing the recent poleward expansion of the Hadley circulation, because of the effect of SST on tropospheric temperature, tropopause height, and baroclinic wave position and stability[45,46]. Further assessments of marine and terrestrial palaeoclimate networks, including compilations of hydroclimate-sensitive records currently in development, have the potential to provide a context for the role of rapid tropical ocean warming during the industrial era in widening the tropical climate belt.

Differences between Southern Hemisphere reconstructions and simulations, and the lack of suitable palaeoclimate records from the extratropical Southern Hemisphere oceans, currently preclude an assessment of the role of the oceans in the delayed onset of Antarctic warming that is seen in regional (Fig. 2) and site-level (Fig. 4a) palaeoclimate analyses, and in observational records[26]. Idealized model experiments of ocean heating predict a centennial-scale delayed onset of anthropogenic warming in the Southern Ocean caused by upwelling of unmodified subsurface water and northward advection of any surface warming signal[11,47]. The strengthening of westerly winds over the Southern Ocean during the twentieth century, related to the Southern Annular Mode[48], have probably also influenced the delayed development of sustained industrial-era warming over Antarctica. In the Northern Hemisphere, the scarcity of suitable marine palaeoclimate records from the mid- to high latitudes also precludes an assessment of ocean–land relationships during the onset of industrial-era warming. However, we find that in full-forcing and greenhouse-only simulations the onset of sustained surface ocean warming in the North Atlantic Ocean is delayed, or instead is characterized by cooling (Extended Data Fig. 8). This finding is consistent with reports of an unusual slowdown of Atlantic Meridional Overturning Circulation during the twentieth century[49]. Increasing knowledge of the temperature evolution of the extratropical oceans before and during the industrial era should be considered a key target for future palaeoclimate research.

The spatial development of industrial-era warming across the oceans and continents demonstrates that the tropical oceans and Northern Hemisphere were particularly responsive to the climate forcings that shaped industrial-era warming. The mid-nineteenth-century commencement of industrial-era warming suggests that Earth's surface temperature may respond to even small increases in greenhouse gas

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

forcing more rapidly than previously thought[50], and highlights the importance of multi-century palaeoclimate records and model simulations in assessing the response of worldwide climate to anthropogenic greenhouse gas emissions.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 9 June 2015; accepted 28 June 2016.

1. Hegerl, G. C., Crowley, T. J., Hyde, W. T. & Frame, D. J. Climate sensitivity constrained by temperature reconstructions over the past seven centuries. *Nature* **440,** 1029–1032 (2006).
2. Mann, M. E. *et al.* Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia. *Proc. Natl Acad. Sci. USA* **105,** 13252–13257 (2008).
3. Frank, D. C. *et al.* Ensemble reconstruction constraints on the global carbon cycle sensitivity to climate. *Nature* **463,** 527–530 (2010).
4. PAGES 2k Consortium. Continental-scale temperature variability during the past two millennia. *Nat. Geosci.* **6,** 339–346 (2013).
5. Neukom, R. *et al.* Inter-hemispheric temperature variability over the past millennium. *Nat. Clim. Chang.* **4,** 362–367 (2014).
6. Levitus, S. *et al.* World ocean heat content and thermosteric sea level change (0–2000 m), 1955–2010. *Geophys. Res. Lett.* **39,** L10603 (2012).
7. England, M. H. *et al.* Recent intensification of wind-driven circulation in the Pacific and the ongoing warming hiatus. *Nat. Clim. Chang.* **4,** 222–227 (2014).
8. Chen, X. & Tung, K.-K. Varying planetary heat sink led to global-warming slowdown and acceleration. *Science* **345,** 897–903 (2014).
9. Karl, T. R. *et al.* Possible artifacts of data biases in the recent global surface warming hiatus. *Science* **348,** 1469–1472 (2015).
10. Crowley, T. J., Obrochta, S. P. & Liu, J. Recent global temperature "plateau" in the context of a new proxy reconstruction. *Earths Futur.* **2,** 281–294 (2014).
11. Marshall, J. *et al.* The ocean's role in the transient response of climate to abrupt greenhouse gas forcing. *Clim. Dyn.* **44,** 2287–2299 (2015).
12. Lewis, S. L. & Maslin, M. A. Defining the Anthropocene. *Nature* **519,** 171–180 (2015).
13. Ruddiman, W. F. The Anthropocene. *Annu. Rev. Earth Planet. Sci.* **41,** 45–68 (2013).
14. Tierney, J. E. *et al.* Tropical sea-surface temperatures for the past four centuries reconstructed from coral archives. *Paleoceanography* **30,** 226–252 (2015).
15. McGregor, H. V. *et al.* Robust global ocean cooling trend for the pre-industrial Common Era. *Nat. Geosci.* **8,** 671–677 (2015).
16. McKay, N. P. & Kaufman, D. S. An extended Arctic proxy temperature database for the past 2,000 years. *Sci. Data* **1,** 140026 (2014).
17. Schmidt, G. A. *et al.* Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1.0). *Geosci. Model Dev.* **4,** 33–45 (2011).
18. Crespin, E., Goosse, H., Fichefet, T., Mairesse, A. & Sallaz-Damaz, Y. Arctic climate over the past millennium: annual and seasonal responses to external forcings. *Holocene* **23,** 321–329 (2013).
19. Otto-Bliesner, B. L. *et al.* Climate variability and change since 850 CE: an ensemble approach with the Community Earth System Model (CESM). *Bull. Am. Meteorol. Soc.* **97,** 735–754 (2015).
20. Schurer, A. P., Tett, S. F. B. & Hegerl, G. C. Small influence of solar variability on climate over the past millennium. *Nat. Geosci.* **7,** 104–108 (2014).
21. Phipps, S. J. *et al.* Paleoclimate data–model comparison and the role of climate forcings over the past 1500 years. *J. Clim.* **26,** 6915–6936 (2013).
22. Sigl, M. *et al.* Timing and climate forcing of volcanic eruptions for the past 2,500 years. *Nature* **523,** 543–549 (2015).
23. Abram, N. J. *et al.* Acceleration of snow melt in an Antarctic Peninsula ice core during the twentieth century. *Nat. Geosci.* **6,** 404–411 (2013).
24. Hannig, J. & Marron, J. S. Advanced distribution theory for SiZer. *J. Am. Stat. Assoc.* **101,** 484–499 (2006).
25. Steig, E. J. *et al.* Warming of the Antarctic ice-sheet surface since the 1957 International Geophysical Year. *Nature* **457,** 459–462 (2009).
26. Jones, J. M. *et al.* Assessing recent trends in high-latitude Southern Hemisphere surface climate. *Nat. Clim. Chang.* (in the press).
27. Screen, J. A. & Simmonds, I. The central role of diminishing sea ice in recent Arctic temperature amplification. *Nature* **464,** 1334–1337 (2010).
28. Hawkins, E. & Sutton, R. Time of emergence of climate signals. *Geophys. Res. Lett.* **39,** L01702 (2012).
29. Diffenbaugh, N. & Scherer, M. Observational and model evidence of global emergence of permanent, unprecedented heat in the 20th and 21st centuries. *Clim. Change* **107,** 615–624 (2011).
30. Mora, C. *et al.* The projected timing of climate departure from recent variability. *Nature* **502,** 183–187 (2013).
31. Lyu, K., Zhang, X., Church, J. A., Slangen, A. B. A. & Hu, J. Time of emergence for regional sea-level change. *Nat. Clim. Chang.* **4,** 1006–1010 (2014).
32. PAGES 2k-PMIP3 group. Continental-scale temperature variability in PMIP3 simulations and PAGES 2k regional temperature reconstructions over the past millennium. *Clim. Past* **11,** 1673–1699 (2015).
33. Bracegirdle, T. J., Stephenson, D. B., Turner, J. & Phillips, T. The importance of sea ice area biases in 21st century multimodel projections of Antarctic temperature and precipitation. *Geophys. Res. Lett.* **42,** 10832–10839 (2015).
34. Schurer, A. P., Hegerl, G. C. & Obrochta, S. P. Determining the likelihood of pauses and surges in global warming. *Geophys. Res. Lett.* **42,** 5974–5982 (2015).
35. Stott, P. A. *et al.* Detection and attribution of climate change: a regional perspective. *WIREs Clim. Chang.* **1,** 192–211 (2010).
36. Bakun, A. Global climate change and intensification of coastal ocean upwelling. *Science* **247,** 198–201 (1990).
37. McGregor, H. V., Dima, M., Fischer, H. W. & Mulitza, S. Rapid 20th-century increase in coastal upwelling off northwest Africa. *Science* **315,** 637–639 (2007).
38. Narayan, N., Paul, A., Mulitza, S. & Schulz, M. Trends in coastal upwelling intensity during the late 20th century. *Ocean Sci.* **6,** 815–823 (2010).
39. Liu, Z., Vavrus, S., He, F., Wen, N. & Zhong, Y. Rethinking tropical ocean response to global warming: the enhanced equatorial warming. *J. Clim.* **18,** 4684–4700 (2005).
40. Xie, S.-P. *et al.* Global warming pattern formation: sea surface temperature and rainfall. *J. Clim.* **23,** 966–986 (2010).
41. Collins, M. *et al.* The impact of global warming on the tropical Pacific Ocean and El Niño. *Nat. Geosci.* **3,** 391–397 (2010).
42. Ji, F., Wu, Z., Huang, J. & Chassignet, E. P. Evolution of land surface air temperature trend. *Nat. Clim. Chang.* **4,** 462–466 (2014).
43. Brönnimann, S. *et al.* Southward shift of the northern tropical belt from 1945 to 1980. *Nat. Geosci.* **8,** 969–974 (2015).
44. Seidel, D. J., Fu, Q., Randel, W. J. & Reichler, T. J. Widening of the tropical belt in a changing climate. *Nat. Geosci.* **1,** 21–24 (2008).
45. Hu, Y., Zhou, C. & Liu, J. Observational evidence for poleward expansion of the Hadley circulation. *Adv. Atmos. Sci.* **28,** 33–44 (2011).
46. Staten, P. W., Rutz, J. J., Reichler, T. & Lu, J. Breaking down the tropospheric circulation response by forcing. *Clim. Dyn.* **39,** 2361–2375 (2012).
47. Armour, K. C., Marshall, J., Scott, J. R., Donohoe, A. & Newsom, E. R. Southern Ocean warming delayed by circumpolar upwelling and equatorward transport. *Nat. Geosci.* **9,** 549–554 (2016).
48. Abram, N. J. *et al.* Evolution of the Southern Annular Mode during the past millennium. *Nat. Clim. Chang.* **4,** 564–569 (2014).
49. Rahmstorf, S. *et al.* Exceptional twentieth-century slowdown in Atlantic Ocean overturning circulation. *Nat. Clim. Chang.* **5,** 475–480 (2015).
50. Ricke, K. L. & Caldeira, K. Maximum warming occurs about one decade after a carbon dioxide emission. *Environ. Res. Lett.* **9,** 124002 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the many scientists who made their published palaeoclimate datasets available via public data repositories. This work developed out of the PAGES (Past Global Changes) Ocean2k working group; we are grateful to K. Anchukaitis, H. Wu, C. Giry, D. Oppo and V. Ersek for their contributions to the Ocean2k syntheses, to the more than 75 volunteers who constructed the Ocean2k phase 1 metadatabase[14,15], and to K. Anchukaitis and V. Trouet for discussions. We thank P. Petrelli, F. Klein and A. Schurer for assistance in obtaining model datasets, and K. McGregor for editorial assistance. We acknowledge support from PAGES funded by the US and Swiss National Science Foundations (NSF) and NOAA, and thank T. Kiefer, M.-F. Loutre and the PAGES 2k Network Coordinators for organizational support. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modelling groups for producing and making available their model output. The US Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support for CMIP and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. N.J.A. is supported by an Australian Research Council (ARC) QEII fellowship awarded under DP110101161 and this work contributes to ARC Discovery Project DP140102059 (N.J.A., M.A.J.C.) and the ARC Centre of Excellence for Climate System Science (N.J.A., S.J.P., J.G.). H.V.M. is supported by ARC Future Fellowship FT140100286 and acknowledges funding from ARC Discovery Project DP1092945 (H.M.V., S.J.P.). We acknowledge fellowship support from a CSIC-Ramón y Cajal post-doctoral programme RYC-2013-14073 (B.M.), a Clare Hall College Cambridge Shackleton Fellowship (B.M.), and an ARC DECRA fellowship DE130100668 (J.G.). We acknowledge research support from US NSF grant OCE1536249 (M.N.E.), the ARC Special Research Initiative for the Antarctic Gateway Partnership (Project ID SR140300001; S.J.P.), Red CONSOLIDER GRACCIE CTM2014-59111-REDC (B.M.), Swiss NSF grant PZ00P2_154802 (R.N.), the Danish Council for Independent Research, Natural Science OCEANHEAT project 12-126709/FNU (M.-S.S.), the National Natural Science Foundation of China (41273083; K.S.) and Shanghai Fund (2013SH012; K.S.). This is University of Maryland Center for Environmental Science contribution 5206.

Author Contributions N.J.A. designed the study with input from H.V.M., J.E.T., M.N.E., N.P.M. and D.S.K. The palaeoclimate data and model analysis was led by N.J.A. with assistance provided by R.N., K.T., B.M., H.G., S.J.P. and E.J.S.; R.N. and J.G. produced the terrestrial Australasia 2k reconstruction; N.J.A., J.E.T., M.N.E., K.H.K., C.P.S. and J.Z. contributed expertise on the high-resolution marine database and reconstructions; H.V.M., M.N.E., B.M., K.T., G.L., J.A.A., P.G.M., M.-S.S., M.-A.S., K.S. and H.L.F. contributed expertise on the moderate-resolution

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

marine database; H.G., S.J.P. and N.J.A. contributed expertise on climate model output, and N.P.M., D.S.K., R.N., E.J.S., J.G., M.A.J.C. and L.v.G. contributed expertise on the terrestrial databases and reconstructions. All authors contributed to discussions that shaped the study and the manuscript. N.J.A. led the writing with contributions from all authors.

Author Information The data and code needed to reproduce the results are available at the World Data Service for Paleoclimatology (http://www.ncdc.noaa.gov/paleo/study/20083). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.J.A. (nerilie.abram@anu.edu.au).

Reviewer Information *Nature* thanks Z. Liu, T. Watanabe and the other anonymous reviewer(s) for their contribution to the peer review of this work.

PAGES 2k Consortium
Kaustubh Thirumalai[8], Belen Martrat[9,10], Hugues Goosse[11], Steven J. Phipps[12,13,14], Eric J. Steig[15,16], K. Halimeda Kilbourne[17], Casey P. Saenger[18], Jens Zinke[19,20,21], Guillaume Leduc[22], Jason A. Addison[23], P. Graham Mortyn[24], Marit-Solveig Seidenkrantz[25], Marie-Alexandrine Sicre[26], Kandasamy Selvaraj[27], Helena L. Filipsson[28], Raphael Neukom[29], Joelle Gergis[30,31], Mark A.J. Curran[32] & Lucien von Gunten[33]

[8]Institute for Geophysics, Jackson School of Geosciences, University of Texas, Austin, Texas 78758, USA. [9]Department of Environmental Chemistry, Institute of Environmental Assessment and Water Research (IDÆA), Spanish Council for Scientific Research (CSIC), 08034 Barcelona, Spain. [10]Department of Earth Sciences, Downing Street, University of Cambridge, Cambridge CB2 3EQ, UK. [11]Earth and Life Institute, Université de Louvain, Place pasteur 3, 1348 Louvain-la-Neuve, Belgium. [12]ARC Centre of Excellence for Climate System Science, University of New South Wales, Sydney, New South Wales 2052, Australia. [13]Climate Change Research Centre, University of New South Wales, Sydney, New South Wales 2052, Australia. [14]Institute for Marine and Antarctic Studies, University of Tasmania, Hobart, Tasmania 7001, Australia. [15]School of Geosciences, University of Edinburgh, Edinburgh EH9 3FE, UK. [16]Department of Earth and Space Sciences, University of Washington, Seattle, Washington 98195 USA. [17]Chesapeake Biological Laboratory, University of Maryland Center for Environmental Science, Solomons, Maryland 20688, USA. [18]Joint Institute for the Study of the Atmosphere and Ocean, University of Washington, Seattle, Washington, USA. [19]Curtin University of Technology, Department of Environment and Agriculture, Bentley, Western Australia 6845, Australia. [20]Australian Institute of Marine Science, 39 Fairway, Nedlands, Western Australia 6009, Australia. [21]Institute of Geological Sciences, Section Palaeontology, Freie Universität Berlin, Malteserstrasse 74-100, 12249 Berlin, Germany. [22]Aix Marseille Université, CNRS, IRD, Coll France, CEREGE, 13545 Aix-en-Provence, France. [23]US Geological Survey, 345 Middlefield Road, MS 910, Menlo Park, California 94025, USA. [24]Universitat Autonoma de Barcelona, Institute of Environmental Science and Technology (ICTA) and Department of Geography, Bellaterra 08193, Spain. [25]Centre for Past Climate Studies and Arctic Research Centre, Department of Geoscience, Aarhus University, Hoegh-Guldbergs Gade 2, DK-8000 Aarhus C, Denmark. [26]Sorbonne Universités (UPMC, Université Paris 06)-CNRS-IRD-MNHN, LOCEAN Laboratory, 4 place Jussieu, F-75005 Paris, France. [27]State Key Laboratory of Marine Environmental Science and Department of Geological Oceanography, Xiamen University, Xiamen 361102, China. [28]Department of Geology, Lund University, Sölvegatan 12, SE-223 62 Lund, Sweden. [29]Oeschger Centre for Climate Change Research and Institute of Geography, University of Bern, 3012 Bern, Switzerland. [30]School of Earth Sciences, University of Melbourne, Melbourne, Victoria 3010, Australia. [31]ARC Centre of Excellence for Climate System Science, University of Melbourne, Melbourne, Victoria 3010, Australia. [32]Australian Antarctic Division and Antarctic Climate and Ecosystems Cooperative Research Centre, Hobart, Tasmania 7000, Australia. [33]PAGES International Project Office, Falkenplatz 16, 3012 Bern, Switzerland.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

Palaeoclimate data. We use the newly developed SST reconstructions and palaeoclimate databases of the PAGES Ocean2k working group. For full details of the data selection criteria and in-depth discussions of the marine datasets and reconstructions, see refs 14 and 15.

The high-resolution (annual or better) component of the Ocean2k dataset consists of 57 coral records (including multiple sensors for some records; Extended Data Table 1), which were used to reconstruct regional mean SST histories for different sectors of the tropical oceans[14]. The SST reconstructions for the tropical Indian, western Pacific and western Atlantic regions are statistically robust over most of the past 400 years[14]. The same reconstruction method applied to the tropical eastern Pacific region yielded poorer statistics and a twentieth-century SST trend that is stronger than suggested by instrumental records. This was attributed to nonlinear hydrologic effects in the eastern tropical Pacific[14]; hence, we exclude this regional reconstruction from the current study. We use the 'best' reconstruction for each region, as identified in ref. 14 on the basis of validation statistics from the ensemble of reconstructions produced. From the Ocean2k low-resolution database[15], we use a subset of 21 marine records that are suitable for the recent temperature trend analysis carried out in this study (Extended Data Fig. 7a). These 21 records meet additional criteria of having strong chronological control through ^{210}Pb profiling or counting of annual sediment layers or coral growth bands, as well as an average sample resolution of 25 years or better; we refer to these as moderate-resolution records. The sense of the proxy-temperature response of the high- and moderate-resolution ocean palaeoclimate records is based on known physical relationships for the incorporation of geochemical and biological tracers into these records or on individually assessed and published temperature–growth relationships.

We compare the ocean records to the continental-scale temperature reconstructions and palaeoclimate databases developed in phase 1 of the PAGES 2k project[4], including the updated Arctic v1.1.1 reconstruction and database[16]. For the North America region we use the tree-ring-based reconstruction, which has decadal resolution, rather than the lower-resolution pollen-derived temperature reconstruction. To avoid record duplication, all marine records were removed from the PAGES 2k continental database before site-level trend analysis. We also exclude the four instrumental records in the South American 2k database from our site-level data assessment so that all information is derived solely from palaeoclimate archives. At the level of regional temperature reconstructions, a small degree of overlap exists in the contributing data used for some of the previously published reconstructions. Specifically, 9 of the 28 records used for the Australasia 2k temperature reconstruction are also used for the tropical western Pacific SST reconstruction, and 1 is used for the tropical Indian Ocean SST reconstruction. To avoid any bias introduced by data overlap, we use a terrestrial-only Australasia 2k reconstruction that was produced using the same methodology as the original reconstruction[4], but excludes any of the marine geochemistry records that are used in the Ocean2k high-resolution reconstructions. Details of this terrestrial-only reconstruction, and the reconstruction data and statistics, accompany this paper as Supplementary Data 1. The terrestrial-only reconstruction demonstrates close agreement with the original Australasia 2k reconstruction and none of the interpretations presented here is altered by using the original reconstruction instead of the terrestrial-only version.

Matlab data structures containing the site-level proxy data and regional reconstructions used in this study are archived with the National Centers for Environmental Information (NCEI) World Data Service for Paleoclimatology at http://www.ncdc.noaa.gov/paleo/study/20083.

Palaeoclimate data analysis and uncertainties. The analyses performed in this study use annually resolved, unsmoothed input data. For the moderately resolved marine records (resolution given in Extended Data Fig. 7a) and the North America reconstruction (decadal resolution), pseudo-annual data was produced by performing a nearest-neighbour interpolation to produce stepped datasets that continued values across the entire sampling interval that they represent. Chronological uncertainty in the palaeoclimate records and reconstructions is extremely low for the industrial era. Annual layer chronologies would be expected to be known to within ±2–3 yr back to AD 1800, using conservative estimates and not taking into account the 1815 Tambora eruption that left an unambiguous fixed-time marker in many palaeoclimate archives. Therefore, chronological uncertainty in the industrial era is well within the level of interpretability of our estimates for the onset of sustained, significant warming based on the change-point detection method (see Methods section 'Change-point method testing').

The area weightings used to calculate the average tropical ocean temperature histories in Fig. 1d were based on the surface area of the target reconstruction regions on an Earth ellipsoid. These areas are: Indian Ocean, $25.5 \times 10^6\,km^2$; western Pacific, $26.9 \times 10^6\,km^2$; western Atlantic, $5.1 \times 10^6\,km^2$. The areas of the terrestrial reconstruction regions[4] (Fig. 1c) are: Arctic, $34.4 \times 10^6\,km^2$; Europe, $13.0 \times 10^6\,km^2$; Asia, $31.1 \times 10^6\,km^2$; North America, $12.5 \times 10^6\,km^2$; Australasia, $37.9 \times 10^6\,km^2$; South America, $20.0 \times 10^6\,km^2$; Antarctica, $34.4 \times 10^6\,km^2$.

The seasonality of the temperature signal captured by the reconstructions differs between regions, owing to the availability of site-level palaeoclimate records in each region, some of which capture climate information related to only a specific season (for example, a summer growing season for trees). Detailed discussion on seasonality can be found in refs 4, 14 and 15. To summarize: the tropical ocean reconstructions represent April–March (tropical year) annual averages; the Arctic, North America and Antarctic reconstructions represent annual averages; the Australasia reconstruction represents a September–February half-year (warm season) average; and the Europe, Asia and South America reconstructions represent local summer averages. We do not expect the seasonal differences between the regional reconstructions to affect our interpretations of the onset of sustained, significant warming between different regions. Change-point analysis of climate model simulations produces near-identical regional warming onsets when data are compiled as annual averages across all regions, and as season-specific averages that match the reported seasonality for each corresponding palaeoclimate reconstruction (Extended Data Fig. 4).

The effect that uncertainty in the regional temperature reconstructions may have on estimates of the onset of industrial-era warming was assessed using reconstruction ensembles (Extended Data Fig. 1a). Reconstruction ensembles are available for South America, Australasia and the three tropical ocean regions, and were calculated as part of the original reconstruction process by using different methodological choices based on the proxy network, the temperature calibration interval and/or target dataset. We also extend this analysis to reconstruction ensembles available for Northern Hemisphere[3] and Southern Hemisphere[5] average temperature as an additional test of the apparent Southern Hemisphere delay in the onset of industrial-era warming (Extended Data Fig. 1a, Supplementary Fig. 1). Uncertainty in the onset of industrial-era warming related to reconstruction uncertainty has a 5%–95% range of within −3 yr to +25 yr for the three tropical ocean regions (which each have an average $\pm 2\sigma$ range across reconstruction members of approximately 0.1 °C during the nineteenth and twentieth centuries). The Australasia ensembles have an average $\pm 2\sigma$ range of 0.4 °C during the nineteenth and twentieth centuries, and there is a −35 yr to +46 yr range (5%–95%) in onset estimates determined across this ensemble. The South America ensembles have an average $\pm 2\sigma$ range of 0.6 °C during the nineteenth and twentieth centuries, and a −16 yr to +31 yr range (5%–95%) in the onset of industrial-era warming (Extended Data Fig. 1a).

As a first-order estimate, the ranges of onset timings obtained from regions for which reconstruction ensembles are available may provide guidance on how reconstruction uncertainty affects other regions. The regional reconstructions for the Arctic and Europe have a similar magnitude of uncertainty during the nineteenth and twentieth centuries to the South America reconstruction (average $\pm 2\sigma$ of approximately 0.6 °C). Therefore, uncertainty in the onset of industrial-era warming related to regional reconstruction quality may be similar between these regions. Reconstruction uncertainty is higher for Asia (approximately 0.9 °C for ±2 root-mean-square error) and Antarctica (approximately 1.2 °C for ±2 standard error), and so we may expect a larger uncertainty range in onset estimates related to reconstruction quality for these regions. Reconstruction uncertainty for North America is based on decadal-resolution data, and so is not directly comparable to the annual resolution of the other regional reconstructions.

Sample size. No statistical methods were used to predetermine sample size.

Model output. We compare the regional palaeoclimate reconstructions to a multi-model ensemble of transient last-millennium simulations from AD 850 to AD 1850[51] (Extended Data Table 2), completed as part of the Fifth Coupled Model Intercomparison Project (CMIP5). Historical simulations from the same ensemble were used to extend the model output from AD 1850 to AD 2005. The CMIP5 last-millennium and historical experiments use transient radiative forcings that include orbital, solar, volcanic, greenhouse and ozone parameters, as well as land-use changes[17,52]. All data were accessed from the Earth System Grid Federation, with the exception of the historical portions of the HadCM3 simulation (provided by A. Schurer) and the FGOALS-s2 simulation (provided by T. Zhou and W. Man).

Multiple simulations of the last-millennium run with LOVECLIM[18] were used to examine climate responses to single radiative forcing scenarios (three ensemble members each) and to assess intra-model variability in full forcing simulations (ten ensemble members). We also examine multiple last-millennium simulations of the CSIRO Mk3L coupled climate model run with progressive addition of radiative forcings (three ensemble members each)[21]. The climate response to greenhouse gas forcing alone was further assessed using single forcing experiments of the HadCM3[20] (four ensemble members) and NCAR CESM1[19] (three ensemble members) models.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

For each of the models we examine surface air temperature (tas field in CMIP5 output files) averaged over the PAGES 2k continental palaeoclimate reconstruction regions, and sea surface temperature (tos field in CMIP5 output files) averaged over the PAGES Ocean2k tropical ocean reconstruction regions. Monthly resolution model output was used to generate annual or season-specific averages that correspond to each palaeoclimate reconstruction target region. Because we use the simulations to examine change-points rather than the magnitude of trends, we do not apply any drift corrections to the model output. Change-point analysis was performed on individual model runs and then compiled (Fig. 3, Supplementary Fig. 2, Extended Data Fig. 8), rather than on ensemble averages, to avoid loss of internal variability in the model data.

Matlab data structures containing the model output (compiled as regional time series with annual-average and season-specific resolution) used in this study are archived with the NCEI World Data Service for Paleoclimatology at http://www.ncdc.noaa.gov/paleo/study/20083.

Change-point analysis method. We analyse the trends in the PAGES 2k continental and tropical ocean temperature reconstructions (Fig. 2a, Extended Data Fig. 1a), in the site-level terrestrial and marine palaeoclimate databases (Fig. 4, Extended Data Figs 6, 7) and in model simulations (Fig. 3, Extended Data Figs 2, 8, Supplementary Figs 2, 3) using the SiZer (SIgnificant ZERo crossings of derivatives) method[24]. SiZer determines the sign and significance of trends in time series data across different levels of smoothing using a Gaussian kernel filter. Following the method used in refs 14 and 23, we assess climate change-points from SiZer output by determining the median year of initiation for the most recent significant ($P < 0.1$) and sustained trends across smoothing bandwidths spanning all integer years in the range 15–50 yr. This range of smoothing levels is designed to reduce the influence of interannual-to-decadal climate variability on the detection of a sustained trend, while avoiding shifting the true change-point time if the smoothing window is too long. We assess trends in the time series since AD 1500, and the onset of the most recent significant trend is classified only if the sign of the trend persists through to the most recent end of the record (that is, a sustained trend). Extended Data Fig. 2a shows the SiZer data used to assess the median initiation point for recent significant warming trends across the continents and tropical oceans in palaeoclimate reconstructions (Fig. 2a). Supplementary Fig. 1 shows the SiZer data for multi-model palaeoclimate simulations (Fig. 3a).

Change-point method testing. Change-points determined by the SiZer method were tested on a set of synthetic time series with known warming onset (Extended Data Fig. 3). We also compare SiZer estimates for the synthetic warming onset with change-points determined using linear change-point methods[53,54]. The synthetic time series were designed to test the performance of change-point detection methods across different forms of long-term trends, in the presence of volcanic-style cooling events, and for varying magnitudes and redness (lag-1 autocorrelation) of variability superimposed upon the trend. Each test was carried out across 1,000-member ensembles to generate a distribution of change-point estimates for each method.

Linear change-point detection methods best capture the change-point in synthetic time series when the long-term trend is derived from two straight lines (Extended Data Fig. 3a, series i, ii). However, the SiZer method is more adaptable than linear change-point methods in detecting the true change-point in time series in which the long-term trend is a curve rather than a straight line (Extended Data Fig. 3a, series iii). This is expected to be advantageous for detecting the initial thermodynamic response to increases in atmospheric greenhouse gas levels, which have an accelerating trajectory during the industrial era (Fig. 3d). Previous research has concluded that, despite the complexity of the climate system, there is a near-linear relationship between global radiative forcing changes and the climate response[10,55]. Hence, we would expect industrial era climate trends to be better approximated by a curve than by a simple straight line.

The addition of synthetic volcanic-style cooling events at the time of, and before, the change-point in synthetic series causes only small deviations in SiZer estimates of the climate change-point away from the true synthetic change-point (Extended Data Fig. 3a, series iv–vi). In our tests with a curved (accelerating) warming trend and volcanic-style cooling events centred at 0 yr, −25 yr and −50 yr relative to the onset of the warming trend, the SiZer method returns median times for the warming onset at −1 yr, −9 yr and −13 yr, respectively. This gives us confidence that large volcanic eruptions during the early nineteenth century are not likely to have substantially skewed the detection of the onset of industrial-era warming trends assessed using the SiZer method.

Finally, we test the sensitivity of change-point detection methods to the addition of varying climate variability (autoregressive AR(1) noise) superimposed on the same long-term warming trend. As the magnitude of climate variability increases, the detection of change-points becomes progressively later using the SiZer method (Extended Data Fig. 3b). As a result, climate time series from regions with large interannual-to-multi-decadal variability may have delayed detection of the onset of industrial-era warming relative to regions with small variability, unless the magnitude of the underlying warming trend is also larger in these regions. In our testing, different levels of lag-1 autocorrelation for the AR(1) noise added to an underlying trend does not alter the median estimate for the onset of warming; however, the range of onset estimates about this median becomes greater as autocorrelation increases (Extended Data Fig. 3c).

Across all of the tests examined here, the linear intersection and Bayesian change-point detection methods produce much wider ranges of warming onset estimates than those produced using the SiZer method on the same synthetic ensembles (Extended Data Fig. 3a–c). The linear methods are also less able to detect change-points for trends that are not simple linear functions. On the basis of our change-point method testing, we use the SiZer method in this study, because it appears to be most adaptable and stable for dealing with the climate changes that characterize industrial-era warming.

We apply our method testing to assess the range of uncertainty in estimates of the onset of regional warming related to the SiZer change-point detection method (Extended Data Fig. 1b). This is carried out using an accelerating warming trend upon which AR(1) noise is added that has the same lag-1 autocorrelation and trend-to-variability characteristics as the regional reconstructions. These parameters are estimated by calculating characteristics of residuals about the 15-yr filters (as in Fig. 2a) of the regional reconstructions. Uncertainty in onset estimates related to the SiZer method is small, typically better than ±25 yr (5%–95%). Exceptions are Antarctica, for which the small trend relative to variability does not allow for the detection of significant trends, and Asia, for which strong lag-1 autocorrelation leads to uncertainty in the lower (5%) bound for the onset of warming (Extended Data Fig. 1b).

Emergence testing. We use the regional temperature reconstructions to determine the extent to which industrial-era warming has caused regional climates to emerge above the level of pre-industrial variability. We choose the interval AD 1622–1799 as the climate baseline to test emergence against. The starting point (AD 1622) represents the earliest year for which temperature reconstructions are available for all of the terrestrial and marine regions examined in this study. The end year (AD 1799) was chosen as a time well before the onset of industrial-era warming in any of the regional reconstructions (Table 1). It is also before strong volcanic cooling events associated with the 1809 Unknown and 1815 Tambora eruptions, which could skew the reference period towards cooler states.

Time of emergence is detected when a climate change signal emerges above a defined noise threshold[28]. Here we use a threshold of 2σ of interannual variability above the mean of the reference interval (Fig. 2c). We smooth the reconstructions using filters with widths of 15–50 yr (the same as for our change-point assessments) and calculate the median time when the climate signal (smoothed reconstructions) emerges and stays above the noise threshold. The emergence year is quite insensitive to the level of smoothing we apply to the climate signal across the 15–50-yr filter widths, with the 5%–95% range of emergence estimates being between only 1 yr and 8 yr for the regional reconstructions for which climate emergence is found to occur.

A difference between our climate emergence assessment and previously published time-of-emergence studies is that the palaeoclimate reconstructions allow us to assess climate emergence using a long baseline interval that occurs entirely before the onset of industrial-era warming. Our results demonstrate that, in some regions, industrial-era warming has already caused climate to emerge above the range of natural variability in the AD 1622–1799 reference interval (Table 1). We test the sensitivity of this result to the choice of reference interval and find that all tested reference intervals before the onset of industrial-era warming result in similar regional emergence patterns and timings, whereas reference periods that include parts of the industrial-era warming signal result in later estimates for the time of emergence. For example, in the Arctic reconstruction, time-of-emergence estimates based on different reference periods are: AD 1947 (AD 1500–1799 reference), AD 1930 (AD 1600–1799 reference), AD 1930 (AD 1622–1799 reference; Table 1), AD 1938 (AD 1700–1799 reference), AD 1960 (AD 1800–1899 reference) and AD 1978 (AD 1850–1899 reference). The overlapping interval of the regional reconstructions that we use for our time-of-emergence reference (AD 1622–1799) occurs during the time of coolest conditions during the past 2,000 years[4,15]. Using longer palaeoclimate reference intervals that incorporate earlier, warm intervals of the past 2,000 years will alter time-of-emergence results, but it is not currently possible to perform this assessment consistently between regions, owing to the length limitations of currently available tropical ocean temperature reconstructions.

Instrumental data. We plot temperature trends from gridded instrumental datasets in Fig. 1. The surface air temperature datasets are from the Climate Research Unit (CRU) TS3.22 product[56], or the ERAI-f product[57] for surface air temperature over Antarctica, and the sea surface temperature datasets are from

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

the HadiSST product[58]. The CO_2-equivalent record shown in Fig. 3d is derived from ref. 59.

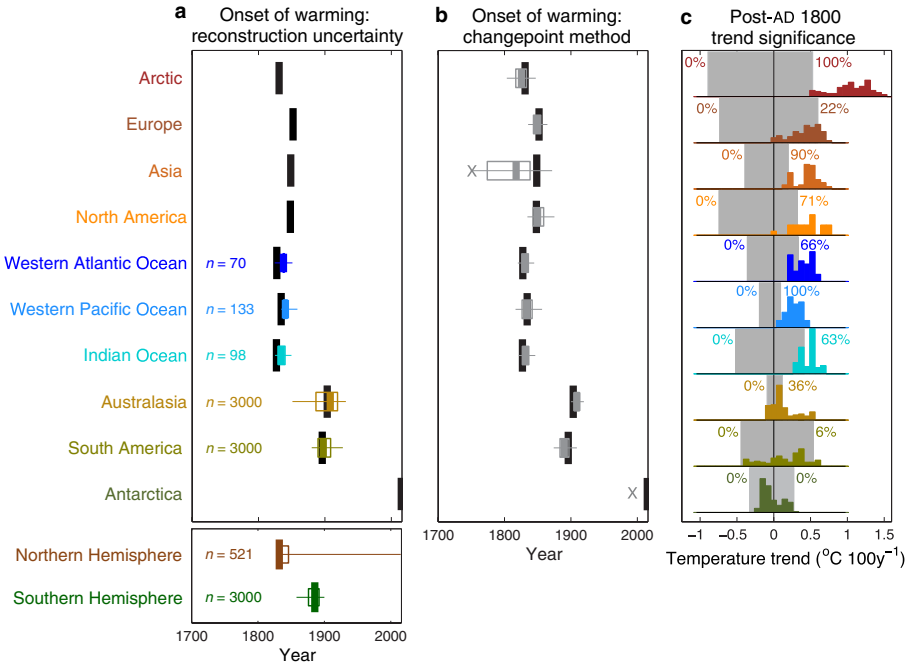
Code availability. Matlab code for assessing the onset (and sign) of industrial-era climate trends using the SiZer method is archived with the NCEI World Data Service for Paleoclimatology at http://www.ncdc.noaa.gov/paleo/study/20083. These files include the original SiZer package[24] obtained from http://www.unc.edu/~marron/marron_software.html, and additional code to assess the onset of sustained, significant temperature trends in annually resolved climate records.

51. Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* **93,** 485–498 (2012).
52. Schmidt, G. A. *et al.* Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1.1). *Geosci. Model Dev.* **5,** 185–191 (2012).
53. Mudelsee, M. Break function regression. *Eur. Phys. J. Spec. Top.* **174,** 49–63 (2009).
54. Ruggieri, E. A Bayesian approach to detecting change points in climatic records. *Int. J. Climatol.* **33,** 520–528 (2013).
55. Crowley, T. J. Causes of climate change over the past 1000 years. *Science* **289,** 270–277 (2000).
56. Harris, I., Jones, P. D., Osborn, T. J. & Lister, D. H. Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 dataset. *Int. J. Climatol.* **34,** 623–642 (2014).
57. Nicolas, J. P. & Bromwich, D. H. New reconstruction of Antarctic near-surface temperatures: multidecadal trends and reliability of global reanalyses. *J. Clim.* **27,** 8070–8093 (2014).
58. Rayner, N. A. *et al.* Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res. Atmos.* **108,** 4407 (2003).
59. MacFarling Meure, C. *et al.* Law Dome CO_2, CH_4 and N_2O ice core records extended to 2000 years BP. *Geophys. Res. Lett.* **33,** L14810 (2006).
60. Oppo, D. W., Rosenthal, Y. & Linsley, B. K. 2,000-year-long temperature and hydrology reconstructions from the Indo-Pacific warm pool. *Nature* **460,** 1113–1116 (2009).
61. Pahnke, K., Sachs, J. P., Keigwin, L., Timmermann, A. & Xie, S.-P. Eastern tropical Pacific hydrologic changes during the past 27,000 years from D/H ratios in alkenones. *Paleoceanography* **22,** PA4214 (2007).
62. Hendy, E. J. *et al.* Abrupt decrease in tropical Pacific sea surface salinity at end of Little Ice Age. *Science* **295,** 1511–1514 (2002).
63. Hendy, I. L., Dunn, L., Schimmelmann, A. & Pak, D. K. Resolving varve and radiocarbon chronology differences during the last 2000 years in the Santa Barbara Basin sedimentary record, California. *Quat. Int.* **310,** 155–168 (2013).
64. Schimmelmann, A., Hendy, I. L., Dunn, L., Pak, D. K. & Lange, C. B. Revised ~2000-year chronostratigraphy of partially varved marine sediment in Santa Barbara Basin, California. *GFF* **135,** 258–264 (2013).
65. Zhao, M., Eglinton, G., Read, G. & Schimmelmann, A. An alkenone ($U_{37}^{K'}$) quasi-annual sea surface temperature record (A.D. 1440 to 1940) using varved sediments from the Santa Barbara Basin. *Org. Geochem.* **31,** 903–917 (2000).
66. Gutiérrez, D. *et al.* Coastal cooling and increased productivity in the main upwelling zone off Peru since the mid-twentieth century. *Geophys. Res. Lett.* **38,** L07603 (2011).
67. Goni, M. A., Thunell, R. C., Woodwort, M. P. & Müller-Karger, F. E. Changes in wind-driven upwelling during the last three centuries: interocean teleconnections. *Geophys. Res. Lett.* **33,** L15604 (2006).
68. Doose-Rolinski, H., Rogalla, U. R., Scheeder, G., Lückge, A. & von Rad, U. High-resolution temperature and evaporation changes during the late Holocene in the northeastern Arabian Sea. *Paleoceanography* **16,** 358–367 (2001).
69. Black, D. E. *et al.* An 8-century tropical Atlantic SST record from the Cariaco Basin: baseline variability, twentieth-century warming, and Atlantic hurricane frequency. *Paleoceanography* **22,** PA4204 (2007).
70. Kuhnert, H. & Mulitza, S. Multidecadal variability and late medieval cooling of near-coastal sea surface temperatures in the eastern tropical North Atlantic. *Paleoceanography* **26,** PA4224 (2011).
71. Leduc, G., Herbert, C. T., Blanz, T., Martinez, P. & Schneider, R. Contrasting evolution of sea surface temperature in the Benguela upwelling system under natural and anthropogenic climate forcings. *Geophys. Res. Lett.* **37,** L20705 (2010).
72. Richter, T. O., Peeters, F. J. C. & van Weering, T. C. E. Late Holocene (0–2.4 ka BP) surface water temperature and salinity variability, Feni Drift, NE Atlantic Ocean. *Quat. Sci. Rev.* **28,** 1941–1955 (2009).
73. Sicre, M. A. *et al.* Sea surface temperature variability in the subpolar Atlantic over the last two millennia. *Paleoceanography* **26,** PA4218 (2011).
74. Abrantes, F. *et al.* Shallow-marine sediment cores record climate variability and earthquake activity off Lisbon (Portugal) for the last 2000 years. *Quat. Sci. Rev.* **24,** 2477–2494 (2005).
75. Versteegh, G. J. M., de Leeuw, J. W., Taricco, C. & Romero, A. Temperature and productivity influences on $U_{37}^{K'}$ and their possible relation to solar forcing of the Mediterranean winter. *Geochem. Geophys. Geosyst.* **8,** Q09005 (2007).
76. Moreno, A. *et al.* The Medieval Climate Anomaly in the Iberian Peninsula reconstructed from marine and lake records. *Quat. Sci. Rev.* **43,** 16–32 (2012).
77. Nieto-Moreno, V. *et al.* Climate conditions in the westernmost Mediterranean over the last two millennia: an integrated biomarker approach. *Org. Geochem.* **55,** 1–10 (2013).
78. Bonnet, S., de Vernal, A., Hillaire-Marcel, C., Radi, T. & Husum, K. Variability of sea-surface temperature and sea-ice cover in the Fram Strait over the last two millennia. *Mar. Micropaleontol.* **74,** 59–74 (2010).
79. Felis, T. *et al.* Subtropical coral reveals abrupt early-twentieth-century freshening in the western North Pacific Ocean. *Geology* **37,** 527–530 (2009).
80. Asami, R. *et al.* Interannual and decadal variability of the western Pacific sea surface condition for the years 1787–2000: reconstruction based on stable isotope record from a Guam coral. *J. Geophys. Res. Oceans* **110,** C05018 (2005).
81. Charles, C. D., Cobb, K., Moore, M. D. & Fairbanks, R. G. Monsoon–tropical ocean interaction in a network of coral records spanning the 20th century. *Mar. Geol.* **201,** 207–222 (2003).
82. Boiseau, M., Ghil, M. & Juillet-Leclerc, A. Climatic trends and interdecadal variability from South-Central Pacific coral records. *Geophys. Res. Lett.* **26,** 2881–2884 (1999).
83. Boiseau, M. *et al.* Atmospheric and oceanic evidences of El Niño-Southern Oscillation events in the south central Pacific Ocean from coral stable isotopic records over the last 137 years. *Paleoceanography* **13,** 671–685 (1998).
84. Tudhope, A. W. *et al.* Variability in the El Niño-Southern Oscillation through a glacial-interglacial cycle. *Science* **291,** 1511–1517 (2001).
85. Alibert, C. & Kinsley, L. A 170-year Sr/Ca and Ba/Ca coral record from the Western Pacific Warm Pool: 1. What can we learn from an unusual coral record? *J. Geophys. Res. Oceans* **113,** C04008 (2008).
86. Quinn, T. M., Taylor, F. W. & Crowley, T. J. Coral-based climate variability in the Western Pacific Warm Pool since 1867. *J. Geophys. Res. Oceans* **111,** C11006 (2006).
87. Linsley, B. K. *et al.* Tracking the extent of the South Pacific Convergence Zone since the early 1600s. *Geochem. Geophys. Geosyst.* **7,** Q05003 (2006).
88. Quinn, T. M., Crowley, T. J. & Taylor, F. W. New stable isotope results from a 173-year coral from Espiritu Santo, Vanuatu. *Geophys. Res. Lett.* **23,** 3413–3416 (1996).
89. Gorman, M. K. *et al.* A coral-based reconstruction of sea surface salinity at Sabine Bank, Vanuatu from 1842 to 2007 CE. *Paleoceanography* **27,** PA3226 (2012).
90. DeLong, K. L., Quinn, T. M., Taylor, F. W., Lin, K. & Shen, C.-C. Sea surface temperature variability in the southwest tropical Pacific since ad 1649. *Nat. Clim. Chang.* **2,** 799–804 (2012).
91. DeLong, K. L., Quinn, T. M., Taylor, F. W., Shen, C.-C. & Lin, K. Improving coral-base paleoclimate reconstructions by replicating 350 years of coral Sr/Ca variations. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **373,** 6–24 (2013).
92. Quinn, T. M. *et al.* A multicentury stable isotope record from a New Caledonia coral: interannual and decadal sea surface temperature variability in the southwest Pacific since 1657 AD. *Paleoceanography* **13,** 412–426 (1998).
93. Bagnato, S., Linsley, B. K., Howe, S. S. & Wellington, G. M. Coral oxygen isotope records of interdecadal climate variations in the South Pacific Convergence Zone region. *Geochem. Geophys. Geosyst.* **6,** Q06001 (2005).
94. Kuhnert, H., Patzold, J., Wyrwoll, K. H. & Wefer, G. Monitoring climate variability over the past 116 years in coral oxygen isotopes from Ningaloo Reef, Western Australia. *Int. J. Earth Sci.* **88,** 725–732 (2000).
95. Druffel, E. R. M. & Griffin, S. Variability of surface ocean radiocarbon and stable isotopes in the southwestern Pacific. *J. Geophys. Res. Oceans* **104,** 23607–23613 (1999).
96. Wu, H. C., Linsley, B. K., Dassie, E. P., Schiraldi, B. & deMenocal, P. B. Oceanographic variability in the South Pacific Convergence Zone region over the last 210 years from multi-site coral Sr/Ca records. *Geochem. Geophys. Geosyst.* **14,** 1435–1453 (2013).
97. Linsley, B. K., Ren, L., Dunbar, R. B. & Howe, S. S. El Niño Southern Oscillation (ENSO) and decadal-scale climate variability at 10°N in the eastern Pacific from 1893 to 1994: a coral-based reconstruction from Clipperton Atoll. *Paleoceanography* **15,** 322–335 (2000).
98. Dunbar, R. B., Wellington, G. M., Colgan, M. W. & Glynn, P. W. Eastern Pacific sea-surface temperature since 1600 A.D.: the $\delta^{18}O$ record of climate variability in Galápagos Corals. *Paleoceanography* **9,** 291–315 (1994).
99. Urban, F. E., Cole, J. E. & Overpeck, J. T. Influence of mean climate change on climate variability from a 155-year tropical Pacific coral record. *Nature* **407,** 989–993 (2000).
100. Cole, J. E. & Fairbanks, R. G. The Southern Oscillation recorded in the $\delta^{18}O$ of corals from Tarawa Atoll. *Paleoceanography* **5,** 669–683 (1990).
101. Cole, J. E., Fairbanks, R. G. & Shen, G. T. Recent variability in the Southern Oscillation: isotopic results from a Tarawa Atoll coral. *Science* **260,** 1790–1793 (1993).
102. Guilderson, T. P. & Schrag, D. P. Reliability of coral isotope records from the Western Pacific Warm Pool: a comparison using age-optimized records. *Paleoceanography* **14,** 457–464 (1999).
103. Linsley, B. K., Dunbar, R. B., Wellington, G. M. & Mucciarone, D. A. A coral-based reconstruction of Intertropical Convergence Zone variability over Central America since 1707. *J. Geophys. Res. Oceans* **99,** 9977–9994 (1994).
104. Cobb, K. M., Charles, C. D., Cheng, H. & Edwards, R. L. El Niño/Southern Oscillation and tropical Pacific climate during the last millennium. *Nature* **424,** 271–276 (2003).
105. Nurhati, I. S., Cobb, K. M., Charles, C. D. & Dunbar, R. B. Late 20th century warming and freshening in the central tropical Pacific. *Geophys. Res. Lett.* **36,** L21606 (2009).
106. Nurhati, I. S., Cobb, K. M. & Di Lorenzo, E. Decadal-scale SST and salinity variations in the central tropical Pacific: signatures of natural and anthropogenic climate change. *J. Clim.* **24,** 3294–3308 (2011).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

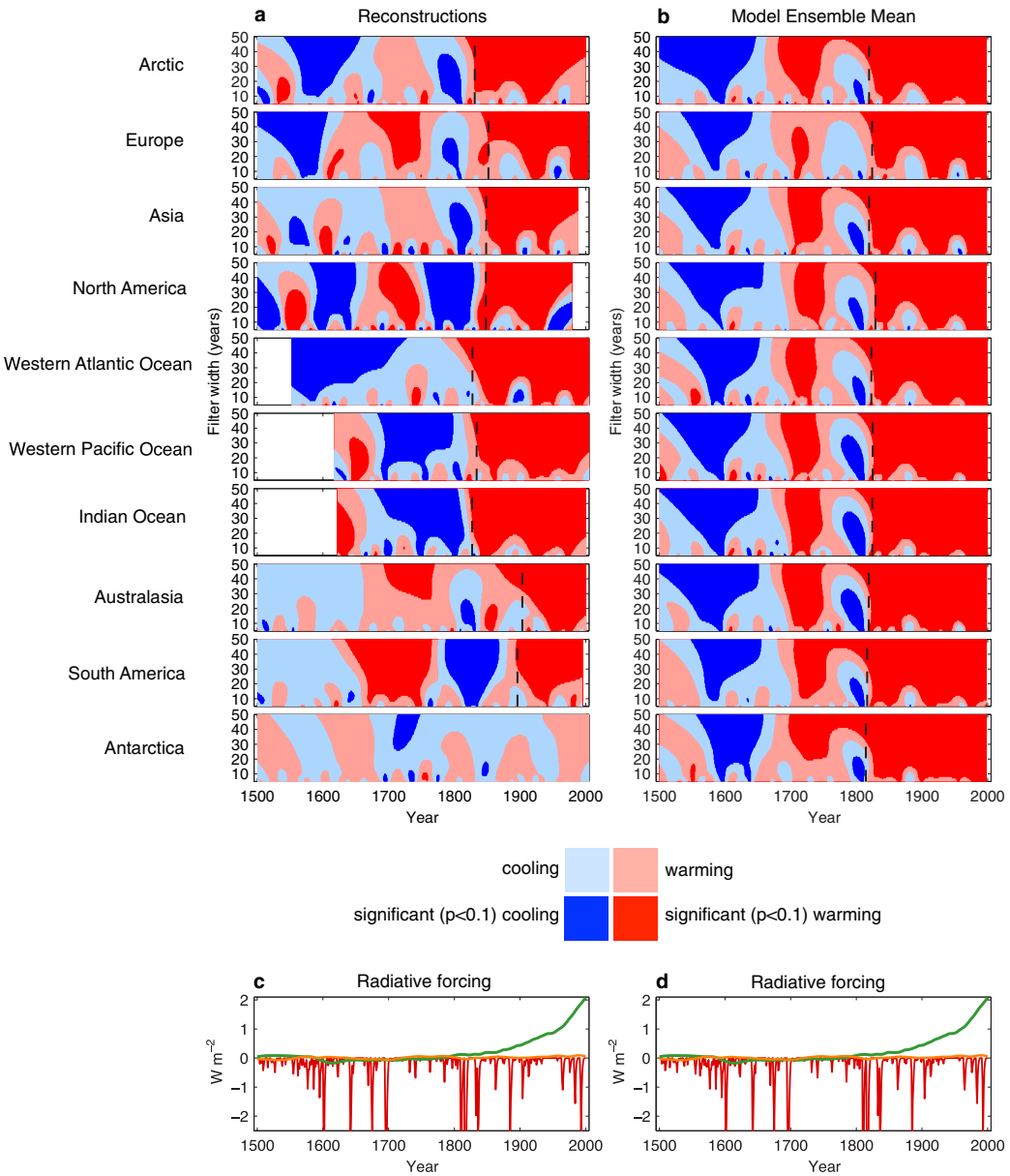
107. Cole, J. E., Dunbar, R. B., McClanahan, T. R. & Muthiga, N. A. Tropical Pacific forcing of decadal SST variability in the western Indian Ocean over the past two centuries. *Science* **287,** 617–619 (2000).
108. Zinke, J., Pfeiffer, M., Timm, O., Dullo, W. C. & Brummer, G. J. A. Western Indian Ocean marine and terrestrial records of climate variability: a review and new concepts on land-ocean interactions since AD 1660. *Int. J. Earth Sci.* **98,** 115–133 (2009).
109. Zinke, J. *et al.* Mayotte coral reveals hydrological changes in the western Indian Ocean between 1881 and 1994. *Geophys. Res. Lett.* **35,** L23707 (2008).
110. Zinke, J., Dullo, W. C., Heiss, G. A. & Eisenhauer, A. ENSO and Indian Ocean subtropical dipole variability is recorded in a coral record off southwest Madagascar for the period 1659 to 1995. *Earth Planet. Sci. Lett.* **228,** 177–194 (2004).
111. Kuhnert, H. *et al.* A 200-year coral stable oxygen isotope record from a high-latitude reef off western Australia. *Coral Reefs* **18,** 1–12 (1999).
112. Damassa, T. D., Cole, J. E., Barnett, H. R., Ault, T. R. & McClanahan, T. R. Enhanced multidecadal climate variability in the seventeenth century from coral isotope records in the western Indian Ocean. *Paleoceanography* **21,** PA2016 (2006).
113. Abram, N. J., Gagan, M. K., Cole, J. E., Hantoro, W. S. & Mudelsee, M. Recent intensification of tropical climate variability in the Indian Ocean. *Nat. Geosci.* **1,** 849–853 (2008).
114. Charles, C. D., Hunter, D. E. & Fairbanks, R. G. Interaction between the ENSO and the Asian monsoon in a coral record of tropical climate. *Science* **277,** 925–928 (1997).
115. Nakamura, N. *et al.* Mode shift in the Indian Ocean climate under global warming stress. *Geophys. Res. Lett.* **36,** L23708 (2009).
116. Pfeiffer, M., Timm, O., Dullo, W. C. & Podlech, S. Oceanic forcing of interannual and multidecadal climate variability in the southwestern Indian Ocean: evidence from a 160 year coral isotopic record (La Reunion, 55°E, 21°S). *Paleoceanography* **19,** PA4006 (2004).
117. Heiss, G. A. *Coral Reefs in the Red Sea: Growth, Production, and Stable Isotopes.* PhD thesis, Christian Albrechts Univ., http://eprints.uni-kiel.de/29262/ (GEOMAR Report 32, GEOMAR Research Center for Marine Geosciences, 1994).
118. Moustafa, Y. *Paleoclimatic Reconstructions of the Northern Red Sea During the Holocene Inferred from Stable Isotope Records of Modern and Fossil Corals and Molluscs.* PhD thesis, Univ. Bremen., http://elib.suub.uni-bremen.de/ip/docs/00010274.pdf (2000).
119. Felis, T. *et al.* A coral oxygen isotope record from the northern Red Sea documenting NAO, ENSO, and North Pacific teleconnections on Middle East climate variability since the year 1750. *Paleoceanography* **15,** 679–694 (2000).
120. Goodkin, N. F., Hughen, K. A., Curry, W. B., Doney, S. C. & Ostermann, D. R. Sea surface temperature and salinity variability at Bermuda during the end of the Little Ice Age. *Paleoceanography* **23,** PA3203 (2008).
121. Hetzinger, S., Pfeiffer, M., Dullo, W.-C., Garbe-Schoenberg, D. & Halfar, J. Rapid 20th century warming in the Caribbean and impact of remote forcing on climate in the northern tropical Atlantic as recorded in a Guadeloupe coral. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **296,** 111–124 (2010).
122. Kilbourne, K. H. *et al.* Paleoclimate proxy perspective on Caribbean climate since the year 1751: evidence of cooler temperatures and multidecadal variability. *Paleoceanography* **23,** PA3220 (2008).
123. Kuhnert, H., Cruger, T. & Patzold, J. NAO signature in a Bermuda coral Sr/Ca record. *Geochem. Geophys. Geosyst.* **6,** Q04004 (2005).
124. Swart, P. K. *et al.* The stable oxygen and carbon isotopic record from a coral growing in Florida Bay: a 160 year record of climatic and anthropogenic influence. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **123,** 219–237 (1996).
125. Swart, P. K., Dodge, R. E. & Hudson, H. J. A 240-year stable oxygen and carbon isotopic record in a coral from South Florida: implications for the prediction of precipitation in Southern Florida. *Palaios* **11,** 362–375 (1996).
126. Saenger, C., Cohen, A. L., Oppo, D. W., Halley, R. B. & Carilli, J. E. Surface-temperature trends and variability in the low-latitude North Atlantic since 1552. *Nat. Geosci.* **2,** 492–495 (2009).
127. Vásquez-Bedoya, L. F., Cohen, A. L., Oppo, D. W. & Blanchon, P. Corals record persistent multidecadal SST variability in the Atlantic Warm Pool since 1775 AD. *Paleoceanography* **27,** PA3231 (2012).
128. DeLong, K. L. *et al.* A reconstruction of sea surface temperature variability in the southeastern Gulf of Mexico from 1734 to 2008 C.E. using cross-dated Sr/Ca records from the coral *Siderastrea siderea*. *Paleoceanography* **29,** 403–422 (2014).
129. Watanabe, S. *et al.* MIROC-ESM 2010: model description and basic results of CMIP5-20c3m experiments. *Geosci. Model Dev.* **4,** 845–872 (2011).
130. Bothe, O., Jungclaus, J. H. & Zanchettin, D. Consistency of the multi-model CMIP5/PMIP3-past1000 ensemble. *Clim. Past* **9,** 2471–2487 (2013).
131. Landrum, L. *et al.* Last millennium climate and its variability in CCSM4. *J. Clim.* **26,** 1085–1111 (2013).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



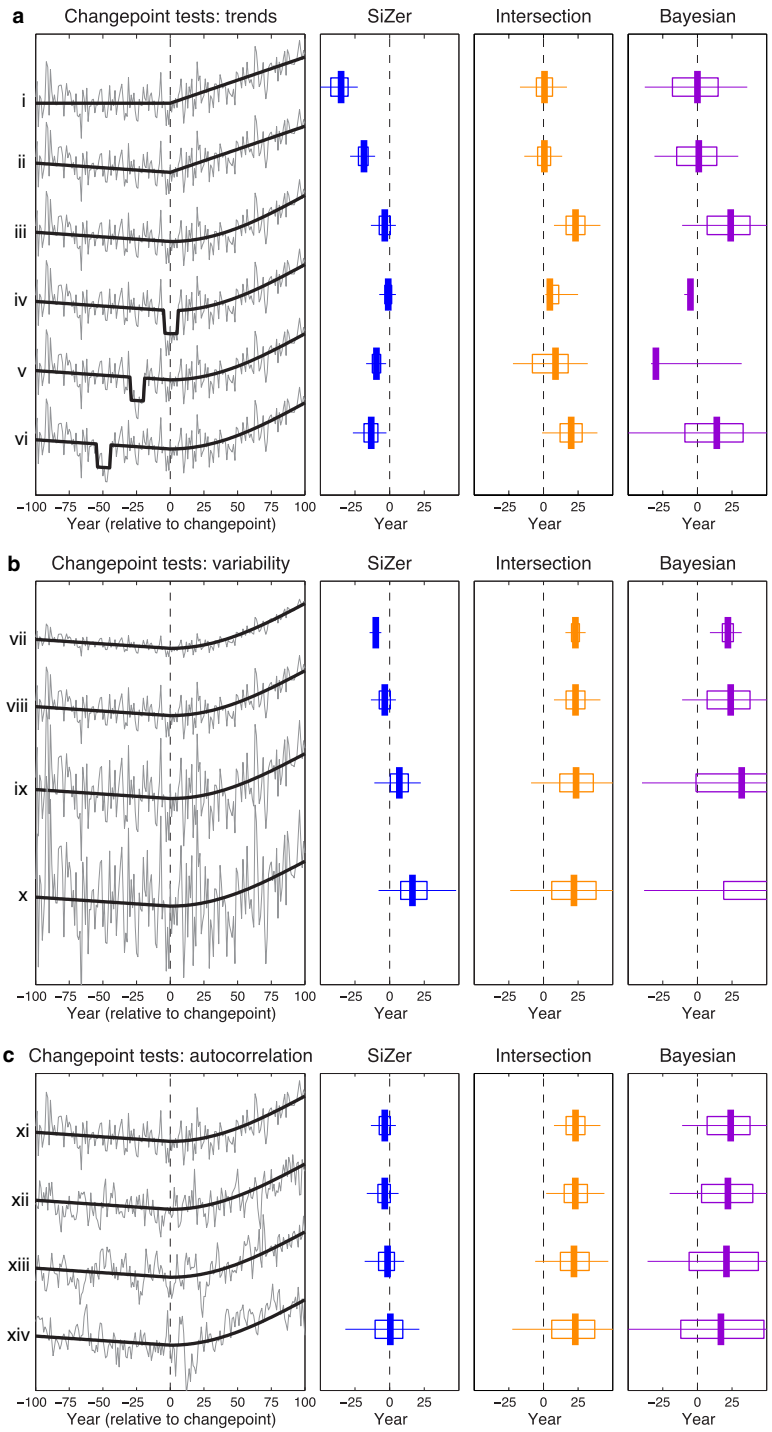
Extended Data Figure 1 | Regional distributions of onset estimates for industrial-era warming and post-AD 1800 warming trends. **a**, Median (black vertical bars; as in Fig. 2a) onset of sustained, significant warming in regional reconstructions, shown with uncertainty ranges in onset estimates based on available reconstruction ensembles (colours; Methods). Distributions of onset estimates related to reconstruction uncertainty denote median (vertical bars), 25%–75% range (boxes) and 5%–95% range (horizontal lines). Distributions for the onset of warming are also calculated for Northern Hemisphere and Southern Hemisphere reconstruction ensembles[3,5], with additional details in Supplementary Fig. 1. The size of reconstruction ensembles is given (*n*). **b**, As in **a**, but for uncertainty ranges in onset estimates based on the SiZer change-point detection method. Distributions (grey) denote the uncertainty in detecting a known warming onset based on synthetic tests, for which 1,000 noise series with lag-1 autocorrelation and trend-to-variability characteristics matching the regional reconstructions are applied to an underlying trend (Methods, Extended Data Fig. 3). Crosses are used for regions for which low trend-to-variability characteristics (Antarctica) or high autocorrelation (Asia) limit the detection of a sustained, significant warming trend in the synthetic tests. **c**, Distribution of century-scale (100-yr) linear warming trends since AD 1800 (coloured bars), shown with reference to the 5%–95% range of century-scale trends beginning during the period AD 1500–1799 (grey shading). Values denote the percentage of century-scale trends since AD 1800 that lie below the 5% level (left) or above the 95% level (right) of trends beginning during the period AD 1500–1799.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



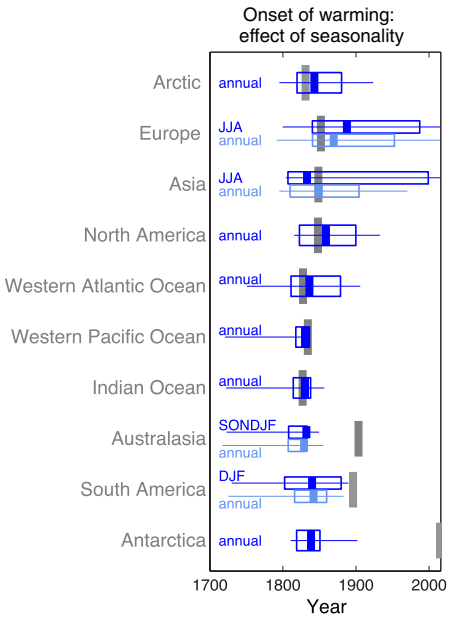
Extended Data Figure 2 | SiZer trend maps used to assess the onset of sustained, significant warming. a, SiZer trend maps[23,24] for each of the regional land and ocean temperature reconstructions. The timing of warming (red) and cooling (blue) trends are calculated at different levels of smoothing. Significant ($P < 0.1$) trends are shown by dark red and dark blue shading. Vertical lines indicate the median onset time for the most recent phase of sustained, significant warming calculated across 15–50-yr filter widths (as used in Fig. 2a; Table 1). **b**, As in **a**, but for the ensemble mean of regional surface air and surface ocean temperatures across CMIP5 last-millennium and historical model simulations. The SiZer analysis for the multi-model ensemble mean is shown for illustrative purposes only and removes the influence of unforced variability to highlight the multi-model thermodynamic response to climate forcings since AD 1500. The multi-model change-point distributions shown in Fig. 3a are based on SiZer analysis of individual experiments (Supplementary Fig. 2), not on the ensemble mean shown here. **c**, **d**, Radiative climate forcings from greenhouse (green), solar (orange) and volcanic (red) sources since AD 1500[17,52]. Note that the magnitude of short-term forcing from large volcanic events[17,22] exceeds the lower limit of the plot axis.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



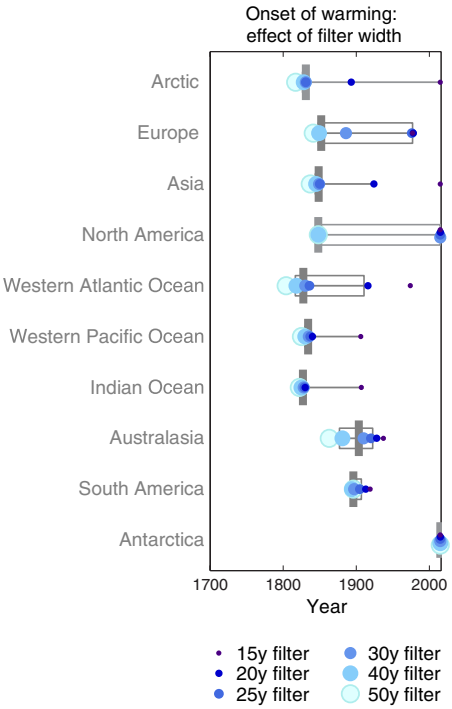
Extended Data Figure 3 | Assessment of change-point detection methods using synthetic time series. a, Example synthetic time series (grey), consisting of long-term trends with a change-point at year 0 (black) and AR(1) noise with a lag-1 autocorrelation of 0.1 and ratio of 100-yr trend to 2σ noise of 1:0.5. The synthetic trends represent: (i) no trend, then linear upward trend; (ii) small linear downward trend, then linear upward trend; (iii) small linear downward trend then accelerating upward trend (one-quarter of a period of a sine curve); (iv–vi) as in (iii), but with 10-year long downward excursions centred at 0 years (iv), −25 years (v) and −50 years (vi) relative to the onset of the accelerating upward trend. Synthetic trends are designed to capture known features of Earth's climate evolution, namely, a long-term gradual pre-industrial cooling trend followed by accelerating industrial-era warming with superimposed episodic volcanic cooling events. The distributions of change-point results using the SiZer method (blue; Methods), the best-fit intersection of two straight lines[53] (orange) and a Bayesian linear change-point method[54] (purple; screening for one change-point and selecting the time of maximum probability) are shown for each experiment. Distributions show the median (thick vertical bars), 25%–75% range (boxes) and 5%–95% range (horizontal lines) of change-points returned across 1,000-member ensembles for each test. **b**, As in **a**, but testing the influence of different magnitudes of AR(1) noise on detecting the onset of warming in an underlying trend (using the small linear downward trend then accelerating upward trend, as in test (iii) in **a**). Tests use a ratio of the 100-yr trend to 2σ noise of 1:0.2 (vii), 1:0.5 (viii), 1:1 (ix) and 1:1.5 (x). **c**, As in **b**, but testing the influence of different AR(1) autocorrelation on detecting the onset of warming in an underlying trend. Tests use lag-1 autocorrelations of 0.1 (xi), 0.3 (xii), 0.5 (xiii) and 0.7 (xiv). See Methods for a detailed discussion of these change-point method tests and Extended Data Fig. 1b for application of the SiZer method tests using signal-to-noise parameters applicable to the regional reconstructions.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



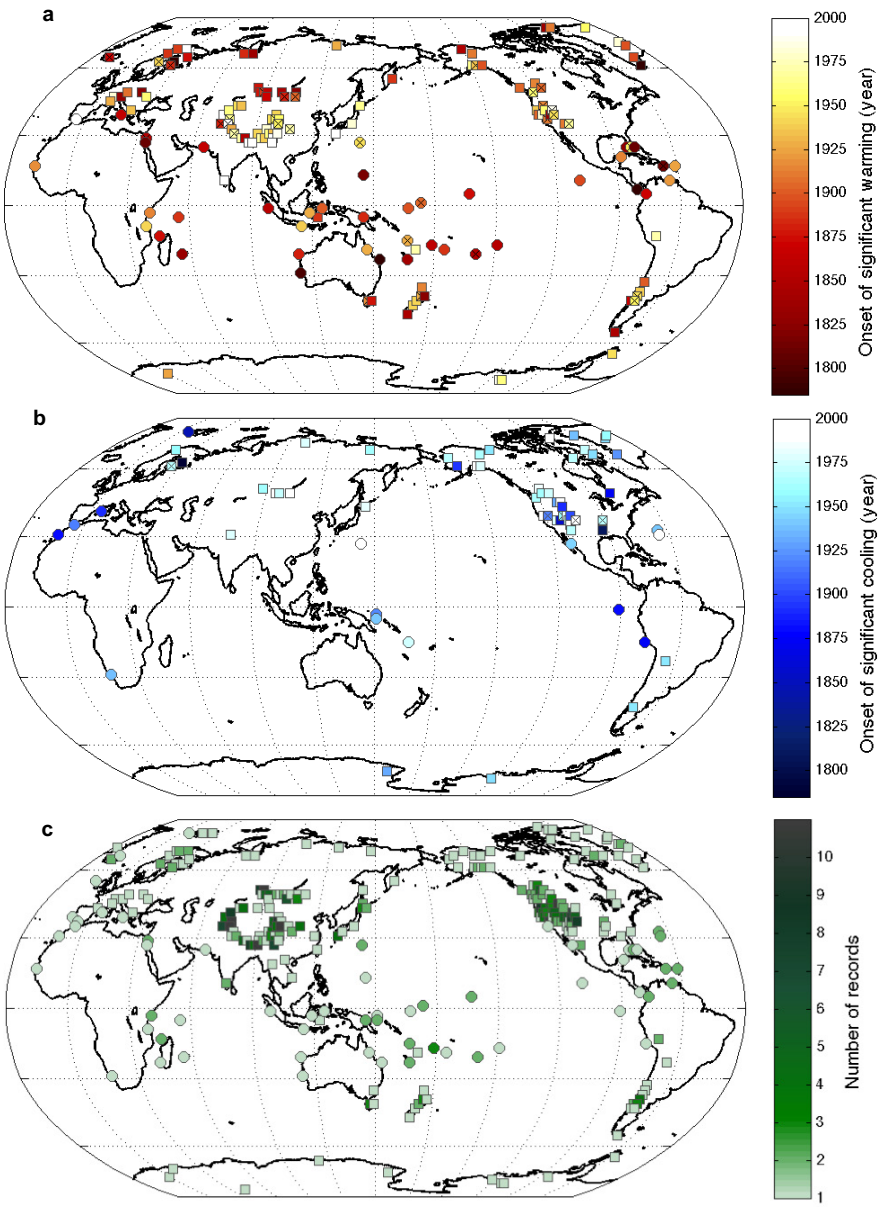
Extended Data Figure 4 | Sensitivity of the onset of sustained, significant warming to seasonality. Median onset of sustained, significant warming in regional reconstructions (grey vertical bars; as in Fig. 2a), shown against the median (blue vertical bars), 25%–75% range (boxes) and 5%–95% range (horizontal lines) of corresponding regional median warming onsets across ten multi-model last-millennium climate simulations with full radiative forcings. Dark blue distribution plots (as in Fig. 3a) show the onset of warming for regions for which regional temperature information from the models has been extracted for annual or season-specific intervals that match the climate representation of the regional reconstructions (as defined in refs 4, 14). For comparison, light blue distributions show the model results for the median onset of sustained, significant warming using annual average data for the reconstruction regions with a seasonal preference (Europe, Asia, Australasia and South America). The similarity between change-point results based on annual average and season-specific model data suggests that it is unlikely that seasonality plays a role in the regional characteristics of the onset of sustained, significant warming described here. JJA, June–August; SONDJF; September–February; DJF, December–February.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 5 | Sensitivity of the onset of sustained, significant warming to filter width. Distributions across 15–50-yr filter widths for the regional onset of sustained, significant warming (Extended Data Fig. 2a) showing the median (grey vertical bars), 25%–75% range (grey boxes) and 5%–95% range (grey horizontal lines). Coloured circles demonstrate the change-points determined at specific filter widths, with markers at AD 2015 indicating that a sustained recent warming is not detected. With the exception of Antarctica, for which significant warming is not observed at any filter width, the change-point analysis shows that shorter filter widths (less smoothing) yield more recent onset dates for sustained warming. This is because decadal-scale variability resets the time over which significant warming is determined to have been sustained in records with less smoothing. This effect accounts for the wide right-side tails produced in warming onset distributions for the Northern Hemisphere reconstruction regions for which decadal-scale variability is particularly strong (Fig. 2).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

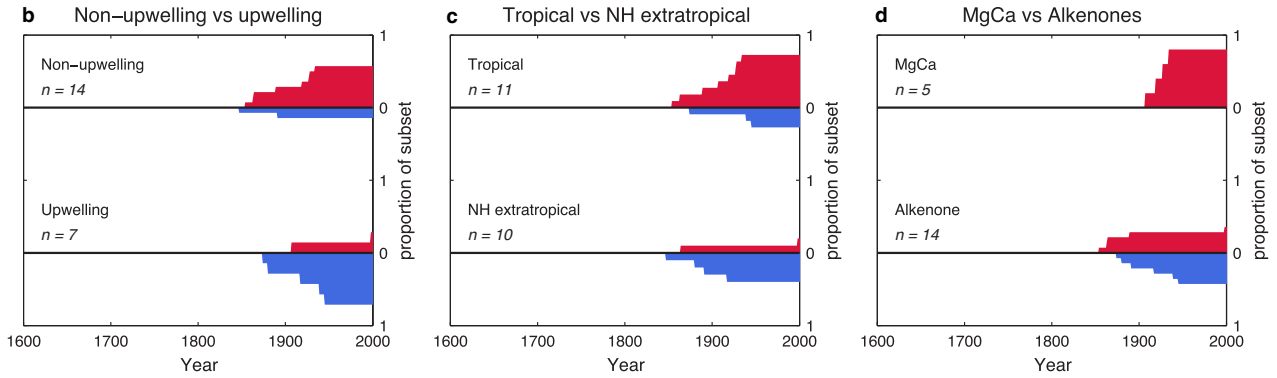


Extended Data Figure 6 | Site-level onset of sustained, significant trends. a, **b**, The same SiZer-based trend analysis performed on the regional reconstructions (Methods, Fig. 2a) was applied to individual temperature-sensitive marine and terrestrial records[4,14–16] to determine the median time of onset of sustained, significant industrial-era warming (**a**) or cooling (**b**) trends. **c**, Number of records available in a 2° × 2° grid region (maximum $n = 28$). Where multiple site-level records are available in a 2° × 2° grid, the median-of-medians change-point is shown in **a** or **b**. Symbols are crossed in **a** and **b** if the interquartile range (25–75 percentile range) of change-points within a grid box exceeds 80 years. In all plots, terrestrial data are shown as squares and marine data are shown as circles. The onset of site-level warming and cooling trends is compiled by latitudinal band in Fig. 4.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

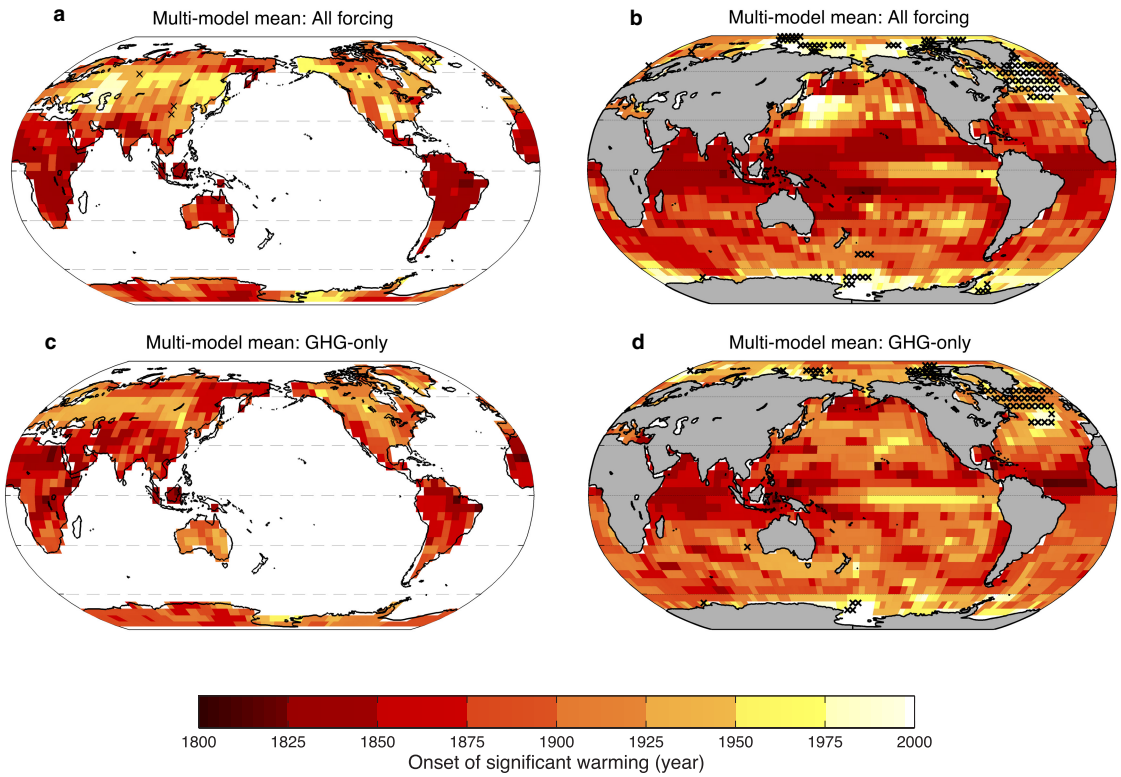
a

Ocean basin	Hemi-sphere	Latitude (°N)	Longitude (°E)	Depth (m)	Mean resolution (max/min) (y sample^{-1})	Seasonality	Upwelling /non-upwelling (*a priori*)	Proxy type	Database number[15]	Ref.
Pacific	Southern	-3.53	119.20	-472	4 (7,<1)	annual	Non	Mg/Ca	0335a	60
Pacific	Northern	4.67	-77.96	-2200	12(22,11)	annual	Non	alkenone	1177	61
Pacific	Northern	4.85	-77.61	-884	10(19,9)	annual	Non	alkenone	1178	61
Pacific	Southern	-18.33	146.45	-10	5(5,5)	annual	Non	coralSr/Ca	1172	62
Pacific	Northern	34.23	-120.02	-590	1(2,<1)	annual	Non	alkenone	1582	63-65
Pacific	Southern	-14.13	-76.50	-299	2(5,1)	annual	Upwelling	alkenone	1571	66
Pacific	Northern	27.90	-111.66	-655	4(9,4)	annual	Upwelling	alkenone	1575	67
Indian	Northern	24.83	65.92	-695	6(15,<1)	annual	Non	alkenone	1574	68
Atlantic	Northern	30.85	-10.10	-355	3(6,<1)	annual	Upwelling	alkenone	0487	37
Atlantic	Northern	10.77	-64.77	-450	1(3,<1)	MAM	Non	Mg/Ca	0039	69
Atlantic	Northern	16.84	-16.73	-323	3(7,0)	JASOND	Non	Mg/Ca	0488	70
Atlantic	Southern	-29.14	16.72	-97	4(20,<1)	annual	Upwelling	alkenone	0484	71
Atlantic	Northern	55.50	-13.90	-2543	13(14,11)	AMJJ	Non	Mg/Ca	0058	72
Atlantic	Northern	66.55	-17.42	-470	3(5,1)	JJA	Non	alkenone	0234	73
Atlantic	Northern	38.56	-9.35	-90	2(6,<1)	ONDJFMA	Non	alkenone	1183	74
Atlantic	Northern	10.65	-64.66	-432	5(11,2)	annual	Upwelling	alkenone	1576	67
Mediterranean	Northern	39.85	17.81	-210	4(4,4)	NDJFM	Non	alkenone	1152	75
Mediterranean	Northern	40.50	4.03	-2394	16(28,2)	annual	Non	alkenone	1157	76
Mediterranean	Northern	35.99	-4.75	-1022	6(7,6)	annual	Upwelling	alkenone	1572a	77
Mediterranean	Northern	36.21	-4.31	-1108	8(12,6)	annual	Upwelling	alkenone	1572b	77
Arctic	Northern	78.92	6.77	-1497	17(18,16)	annual	Non	dinocyst	1147	78



Extended Data Figure 7 | Moderately resolved marine records and their industrial-era trends. a, Details of the moderate-resolution, SST-sensitive marine records[37,60–78] compiled by the Ocean2k working group. See ref. 15 for further details on these records. Acronyms under seasonality refer to months. **b–d**, Cumulative distributions of the onset of sustained, significant warming (red; upward) or cooling (blue; downward) in subsets of the moderately resolved marine records. SiZer-based trend analysis was performed on the site-level records and is compiled using the same subsets examined in the 'mini-bin' analysis of ref. 15. Shading in **a** denotes records for which SiZer analysis detects recent significant cooling trends. The subsets of moderately resolved marine records produce the best differentiation between recent warming and cooling trends when the site-level change-point results are grouped according to non-upwelling (warming) and upwelling (cooling) sites. This suggests that enhanced ocean upwelling during the industrial era may provide a more robust mechanism to account for recent cooling trends in some localized parts of the world oceans, rather than differentiation based on latitude or geochemical analysis type. NH, Northern Hemisphere.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 8 | Onset of sustained, significant warming trends in climate simulations. The same SiZer-based change-point analysis performed on the regional reconstructions (Methods) was applied to climate simulations compiled at 5° × 5° grid resolution to determine the spatial fingerprint for the onset of industrial-era trends. **a**, **b**, The multi-model mean for the onset of sustained, significant warming in surface air temperature (**a**) and sea surface temperature (**b**) calculated for the first ensemble member of LOVECLIM, CSIRO Mk3L, CCSM4 and HadCM3 last-millennium experiments with full transient radiative forcings ('All forcing'). **c**, **d**, As in **a** and **b**, but for the first ensemble member of greenhouse-only ('GHG-only') forced experiments of the same models. Crosses indicate grid boxes in which one or more models produce recent significant cooling rather than warming trends. See Supplementary Fig. 3 for results from individual models. In the subset of models examined here, the fingerprint of warming onset across the oceans is characterized by early warming of most tropical ocean regions except the equatorial eastern Pacific. Delayed warming—or cooling—occurs in the North Atlantic, Arctic and Southern Oceans. The onset of sustained warming is delayed at the grid level in many terrestrial parts of the Northern Hemisphere mid- to high latitudes, owing to large decadal-scale model variability over these regions, a feature that is reduced (but still evident; Extended Data Fig. 5) in regionally averaged climate data. The delayed onset of Antarctic warming is not evident in this grid-level assessment, consistent with Southern Hemisphere data–model disagreement observed in regionally averaged data.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 1 | High-resolution, SST-sensitive coral data compiled by the Ocean2k project

Reconstruction region	Latitude (°N)	Longitude (°E)	Site name	Coral genus	Sampling resolution	Proxy type	References
western Pacific	27.10	142.20	Chichijima	*Porites*	seasonal	Sr/Ca	79
western Pacific	27.10	142.20	Chichijima	*Porites*	seasonal	$\delta^{18}O$	79
western Pacific	13.60	144.83	Double Reef	*Porites*	monthly	$\delta^{18}O$	80
western Pacific	-8.26	115.57	Bali	*Porites*	monthly	$\delta^{18}O$	81
western Pacific	-1.50	124.83	Bunaken	*Porites*	monthly	$\delta^{18}O$	81
western Pacific	-17.50	-149.83	Moorea	*Porites*	annual	$\delta^{18}O$	82,83
western Pacific	-5.22	145.82	Madang	*Porites*	seasonal	$\delta^{18}O$	84
western Pacific	-4.15	144.88	Liang	*Porites*	seasonal	$\delta^{18}O$	84
western Pacific	-2.50	150.5	Kavieng	*Porites*	monthly	Sr/Ca	85
western Pacific	-4.18	151.98	Rabul	*Porites*	monthly	$\delta^{18}O$	86
western Pacific	-4.18	151.98	Rabul	*Porites*	monthly	Sr/Ca	86
western Pacific	-21.24	-159.83	Raratonga*	*Porites*	seasonal	$\delta^{18}O$	87
western Pacific	-21.24	-159.83	Raratonga*	*Porites*	seasonal	Sr/Ca	87
western Pacific	-15.00	166.99	Espiritu Santo	*Porites*	annual	$\delta^{18}O$	88
western Pacific	-15.94	166.04	Sabine Bank*	*Porites*	monthly	$\delta^{18}O$	89
western Pacific	-22.29	166.27	Amedee Island*	*Porites*	monthly	Sr/Ca	90,91
western Pacific	-22.48	166.47	Amadee Island	*Porites*	seasonal	$\delta^{18}O$	92
western Pacific	-16.82	179.23	Suvasuva Bay	*Diploastrea*	annual	$\delta^{18}O$	93
western Pacific	-16.82	179.23	Suvasuva Bay*	*Porites*	annual	$\delta^{18}O$	87
western Pacific	-16.82	179.23	Suvasuva Bay	*Porites*	annual	Sr/Ca	87
western Pacific	-21.91	113.97	Ningaloo Reef	*Porites*	seasonal	$\delta^{18}O$	94
western Pacific	-22.10	153.00	Abraham Reef	*Porites*	annual	$\delta^{18}O$	95
western Pacific	-19.90	-174.70	Tonga	*Porites*	annual	Sr/Ca	96
eastern Pacific	10.28	-109.21	Clipperton Atoll*	*Porites*	seasonal	$\delta^{18}O$	97
eastern Pacific	-0.40	-91.23	Urvina Bay*	*Porites*	annual	$\delta^{18}O$	98
eastern Pacific	1.00	173.00	Maiana Atoll	*Porites*	seasonal	$\delta^{18}O$	99
eastern Pacific	1.00	172.00	Tarawa Atoll	*Porites*	monthly	$\delta^{18}O$	100,101
eastern Pacific	-0.53	166.93	Nauru*	*Porites*	seasonal	$\delta^{18}O$	102
eastern Pacific	7.95	-82.00	Secas Island	*Porites*	seasonal	$\delta^{18}O$	103
eastern Pacific	5.87	-162.13	Palmyra Atoll*	*Porites*	monthly	$\delta^{18}O$	104
eastern Pacific	5.87	-162.13	Palmyra Atoll	*Porites*	monthly	Sr/Ca	105,106
Indian	-3.00	40.00	Malindi	*Porites*	annual	$\delta^{18}O$	107
Indian	-12.65	45.10	Mayotte	*Porites*	seasonal	$\delta^{18}O$	108,109
Indian	-12.65	45.10	Mayotte	*Porites*	seasonal	Sr/Ca	108,109
Indian	-23.15	43.58	Ifaty Reef	*Porites*	seasonal	$\delta^{18}O$	110
Indian	-28.46	113.77	Houtman Abrolhos	*Porites*	seasonal	$\delta^{18}O$	111
Indian	-8.02	39.50	Mafia Island	*Diploastrea*	seasonal	$\delta^{18}O$	112
Indian	-0.13	98.52	Mentawai*	*Porites*	monthly	$\delta^{18}O$	113
Indian	-4.62	55.00	Mahe, Seychelles	*Porites*	monthly	$\delta^{18}O$	114
Indian	-3.20	40.10	Malindi	*Porites*	monthly	$\delta^{18}O$	115
Indian	-21.00	55.00	La Reunion	*Porites*	seasonal	$\delta^{18}O$	116
Indian	29.43	34.97	Aquaba*	*Porites*	annual	$\delta^{18}O$	117
Indian	27.85	34.31	Ras Um Sidd*	*Porites*	seasonal	$\delta^{18}O$	118
Indian	27.85	34.32	Ras Um Sidd	*Porites*	seasonal	$\delta^{18}O$	119
western Atlantic	30.65	-64.99	Bermuda	*Diploria*	annual	$\delta^{18}O$	120
western Atlantic	30.65	-64.99	Bermuda	*Diploria*	annual	Sr/Ca	120
western Atlantic	16.20	-61.49	Guadeloupe	*Diploria*	monthly	$\delta^{18}O$	121
western Atlantic	16.20	-61.49	Guadeloupe	*Diploria*	monthly	Sr/Ca	121
western Atlantic	17.93	-67.00	Puerto Rico	*Montastrea*	annual	$\delta^{18}O$	122
western Atlantic	17.93	-67.00	Puerto Rico	*Montastrea*	annual	Sr/Ca	122
western Atlantic	32.47	-64.70	Bermuda	*Diploria*	monthly	$\delta^{18}O$	123
western Atlantic	32.47	-64.70	Bermuda	*Diploria*	monthly	Sr/Ca	123
western Atlantic	24.93	-80.75	Florida Bay	*Solenastrea*	annual	$\delta^{18}O$	124
western Atlantic	25.38	-80.17	Biscayne, Florida	*Montastrea*	annual	$\delta^{18}O$	125
western Atlantic	25.84	-78.62	Bahamas	*Siderastrea*	annual	growth rate	126
western Atlantic	20.83	-86.74	Yucatan*	*Siderastrea*	annual	growth rate	127
western Atlantic	24.66	-82.83	Dry Tortugas*	*Siderastrea*	monthly	Sr/Ca	128

See refs 79–128 and the database synthesis in ref. 14 for a detailed discussion of the high-resolution coral records. Shading denotes records for which SiZer analysis detects recent significant cooling trends (Extended Data Fig. 6b). Strong multi-decadal variability may account for recent SST cooling signals in site-level coral records from Galapagos, Japan and Bermuda, whereas coral Sr/Ca records from Papua New Guinea are believed to have non-climatic influences.
*Composite of multiple coral records.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 2 | Details of the last-millennium and historical climate simulations

Model	Sea surface temperature (tos files)		Surface air temperature (tas files)		References
	Last Millennium	Historical	Last Millennium	Historical	
BCC-CSM1.1	past1000_r1i1p1	past1000_r1i1p1	past1000_r1i1p1	past1000_r1i1p1	http://forecast.bcccsm.ncc-cma.net /web/channel-43.htm
MIROC-ESM	past1000_r1i1p1	historical_r1i1p1	past1000_r1i1p1	historical_r1i1p1	129
IPSL CM5A-LR	not available	not available	past1000_r1i1p1	historical_r1i1p1	http://icmc.ipsl.fr/index.php/cmip5
MPI-ESM-P	past1000_r1i1p1	historical_r1i1p1	past1000_r1i1p1	thistorical_r1i1p1	130 http://www.mpimet.mpg.de/en/science/models/mpi-esm.html
NCAR CCSM4	past1000_r1i1p1	historical_r1i1p1	past1000_r1i1p1	historical_r1i1p1	131 http://www.cesm.ucar.edu /experiments/
HadCM3	past1000_r1i1p1	pers. comm	past1000_r1i1p1	pers. comm	20
GISS-E2-R*	past1000_r1i1p124	historical_r1i1p124	past1000_r1i1p124	historical_r1i1p124	http://data.giss.nasa.gov/modelE/ar5/
FGOALS-s2	past1000_r1i1p1	pers. comm	not available	not available	
LOVECLIM†	Full forcing (10 member ensemble; e1-e10)		Full forcing (10 member ensemble; e1-e10)		18
LOVECLIM	Greenhouse only (3 member ensemble)		Greenhouse only (3 member ensemble)		18
LOVECLIM	Solar only (3 member ensemble)		Solar only (3 member ensemble)		18
LOVECLIM	Volcanic only (3 member ensemble)		Volcanic only (3 member ensemble)		18
CSIRO-Mk3L v1.2	Orbital + Greenhouse (3 member ensemble)		Orbital + Greenhouse (3 member ensemble)		21
CSIRO-Mk3L v1.2	Orbital + Greenhouse + Solar (3 member ensemble)		Orbital + Greenhouse + Solar (3 member ensemble)		21
CSIRO-Mk3L v1.2†	Orbital + Greenhouse + Solar + Volcanic (3 member ensemble)		Orbital + Greenhouse + Solar + Volcanic (3 member ensemble)		21
HadCM3	Greenhouse only (4 member ensemble)		Greenhouse only (4 member ensemble)		20
NCAR CESM1	Greenhouse only (3 member ensemble)		Greenhouse only (3 member ensemble)		19

This table includes CMIP5 experiment and ensemble member identification and details of additional idealized forcing experiments[18–21,129–131] (see also http://forecast.bcccsm.ncc-cma.net/web/channel-43.htm, http://icmc.ipsl.fr/index.php/cmip5, http://www.mpimet.mpg.de/en/science/models/mpi-esm.html, http://www.cesm.ucar.edu/experiments and http://data.giss.nasa.gov/modelE/ar5). Under CMIP5 protocols[51], last-millennium simulations cover the period AD 850–1850 and historical simulations cover the period AD 1851–2005. The number of models that have run the CMIP5 transient last-millennium experiment is smaller than the full set of CMIP5 climate models.

*Experiment 24 of GISS-E2-R last-millennium ensemble is used for the multi-model assessment (Fig. 3a). The GISS-E2-R ensembles use different combinations of datasets for solar forcing, volcanic forcing and land-use change, but do not include specialized single or cumulative forcing experiments such as those analysed in Fig. 3b–d.

†The first ensemble members of these experiments contribute to the multi-model assessment (Fig. 3a).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

ARTICLE

doi:10.1038/nature19310

Genomic insights into the origin of farming in the ancient Near East

Iosif Lazaridis[1,2], Dani Nadel[3], Gary Rollefson[4], Deborah C. Merrett[5], Nadin Rohland[1], Swapan Mallick[1,2,6], Daniel Fernandes[7,8], Mario Novak[7,9], Beatriz Gamarra[7], Kendra Sirak[7,10], Sarah Connell[7], Kristin Stewardson[1,6], Eadaoin Harney[1,6,11], Qiaomei Fu[1,12,13], Gloria Gonzalez-Fortes[14], Eppie R. Jones[15], Songül Alpaslan Roodenberg[16], György Lengyel[17], Fanny Bocquentin[18], Boris Gasparian[19], Janet M. Monge[20], Michael Gregg[20], Vered Eshed[21], Ahuva-Sivan Mizrahi[21], Christopher Meiklejohn[22], Fokke Gerritsen[23], Luminita Bejenaru[24], Matthias Blüher[25], Archie Campbell[26], Gianpiero Cavalleri[27], David Comas[28], Philippe Froguel[29,30], Edmund Gilbert[27], Shona M. Kerr[26], Peter Kovacs[31], Johannes Krause[32], Darren McGettigan[33], Michael Merrigan[34], D. Andrew Merriwether[35], Seamus O'Reilly[34], Martin B. Richards[36], Ornella Semino[37], Michel Shamoon-Pour[35], Gheorghe Stefanescu[38], Michael Stumvoll[25], Anke Tönjes[25], Antonio Torroni[37], James F. Wilson[39,40], Loic Yengo[29], Nelli A. Hovhannisyan[41], Nick Patterson[2], Ron Pinhasi[7]§ & David Reich[1,2,6]§

We report genome-wide ancient DNA from 44 ancient Near Easterners ranging in time between ~12,000 and 1,400 BC, from Natufian hunter–gatherers to Bronze Age farmers. We show that the earliest populations of the Near East derived around half their ancestry from a 'Basal Eurasian' lineage that had little if any Neanderthal admixture and that separated from other non-African lineages before their separation from each other. The first farmers of the southern Levant (Israel and Jordan) and Zagros Mountains (Iran) were strongly genetically differentiated, and each descended from local hunter–gatherers. By the time of the Bronze Age, these two populations and Anatolian-related farmers had mixed with each other and with the hunter–gatherers of Europe to greatly reduce genetic differentiation. The impact of the Near Eastern farmers extended beyond the Near East: farmers related to those of Anatolia spread westward into Europe; farmers related to those of the Levant spread southward into East Africa; farmers related to those of Iran spread northward into the Eurasian steppe; and people related to both the early farmers of Iran and to the pastoralists of the Eurasian steppe spread eastward into South Asia.

Between 10,000 and 9,000 BC, humans began practicing agriculture in the Near East[1]. In the ensuing five millennia, plants and animals domesticated in the Near East spread throughout West Eurasia (a vast region that also includes Europe) and beyond. The relative homogeneity of present-day West Eurasians in a world context[2] suggests the possibility of extensive migration and admixture that homogenized geographically and genetically disparate sources of ancestry. The spread of the world's first farmers from the Near East would have been a mechanism for such homogenization. To date, however, owing to the poor preservation of DNA in warm climates, it has been impossible to study the population structure and history of the first farmers and to trace their contribution to later populations.

In order to overcome the obstacle of poor DNA preservation, we took advantage of two methodological developments. First, we sampled from the inner ear region of the petrous bone[3,4] which can yield up to ~100 times more endogenous DNA than other skeletal elements[4]. Second, we used in-solution hybridization[5] to enrich extracted DNA for about 1.2 million single nucleotide polymorphism (SNP) targets[6,7], making efficient sequencing practical by filtering out microbial and non-informative human DNA. We merged all sequences extracted from each individual, and randomly sampled a single sequence with minimum mapping and sequence quality to represent each SNP, restricting our investigation to individuals with at least 9,000 SNPs covered at least once (Methods). We obtained genome-wide data that passed

[1]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. [2]Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. [3]The Zinman Institute of Archaeology, University of Haifa, Haifa 3498838, Israel. [4]Department of Anthropology, Whitman College, Walla Walla, Washington 99362, USA. [5]Department of Archaeology, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada. [6]Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts 02115, USA. [7]School of Archaeology and Earth Institute, Belfield, University College Dublin, Dublin 4, Ireland. [8]CIAS, Department of Life Sciences, University of Coimbra, Coimbra 3000-456, Portugal. [9]Institute for Anthropological Research, Zagreb 10000, Croatia. [10]Department of Anthropology, Emory University, Atlanta, Georgia 30322, USA. [11]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138, USA. [12]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig 04103, Germany. [13]Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences, IVPP, CAS, Beijing 100044, China. [14]Department of Biology and Evolution, University of Ferrara, Ferrara I-44121, Italy. [15]Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, UK. [16]J.M. van Nassaulaan 9, Santpoort-Noord 2071 VA, The Netherlands. [17]Department of Prehistory and Archaeology, University of Miskolc, Miskolc-Egyetemváros 3515, Hungary. [18]French National Centre for Scientific Research, UMR 7041, Nanterre Cedex 92023, France. [19]Institute of Archaeology and Ethnology, National Academy of Sciences of the Republic of Armenia, Yerevan 0025, Republic of Armenia. [20]University of Pennsylvania Museum of Archaeology and Anthropology, Philadelphia, Pennsylvania 19104, USA. [21]Israel Antiquities Authority, Jerusalem 91004, Israel. [22]Department of Anthropology, University of Winnipeg, Winnipeg, Manitoba R3B 2E9, Canada. [23]Netherlands Institute in Turkey, Istanbul 34433, Turkey. [24]Faculty of Biology, Alexandru Ioan Cuza University of Iasi, Iasi 700505, Romania. [25]Department of Internal Medicine and Dermatology, Clinic of Endocrinology and Nephrology, Leipzig 04103, Germany. [26]Generation Scotland, Centre for Genomic and Experimental Medicine, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, UK. [27]RCSI Molecular & Cellular Therapeutics, Royal College of Surgeons in Ireland, Dublin 2, Ireland. [28]Institut de Biologia Evolutiva (CSIC-UPF), Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Barcelona 08003, Spain. [29]Univ. Lille, CNRS, Institut Pasteur de Lille, UMR 8199 - EGID, Lille F-59000, France. [30]Imperial College London, Department of Genomics of Common Disease, London Hammersmith Hospital, London W12 0HS, UK. [31]Leipzig University Medical Center, IFB Adiposity Diseases, Leipzig 04103, Germany. [32]Max Planck Institute for the Science of Human History, Jena 07745, Germany. [33]School of History, Newman Building, University College Dublin, Belfield, Dublin 4, Ireland. [34]Genealogical Society of Ireland, Dún Laoghaire, County Dublin, Ireland. [35]Department of Anthropology, Binghamton University, State University of New York, New York 13902, USA. [36]Department of Biological Sciences, School of Applied Sciences, University of Huddersfield, Queensgate, Huddersfield HD1 3DH, UK. [37]Dipartimento di Biologia e Biotecnologie "L. Spallanzani", Università di Pavia, Pavia 27100, Italy. [38]Institutul de Cercetari Biologice, Iaşi 700505, Romania. [39]Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Edinburgh EH8 9AG, UK. [40]MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK. [41]Center of Excellence in Applied Biosciences, Yerevan State University, Yerevan 0025, Republic of Armenia. §These authors jointly supervised this work.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

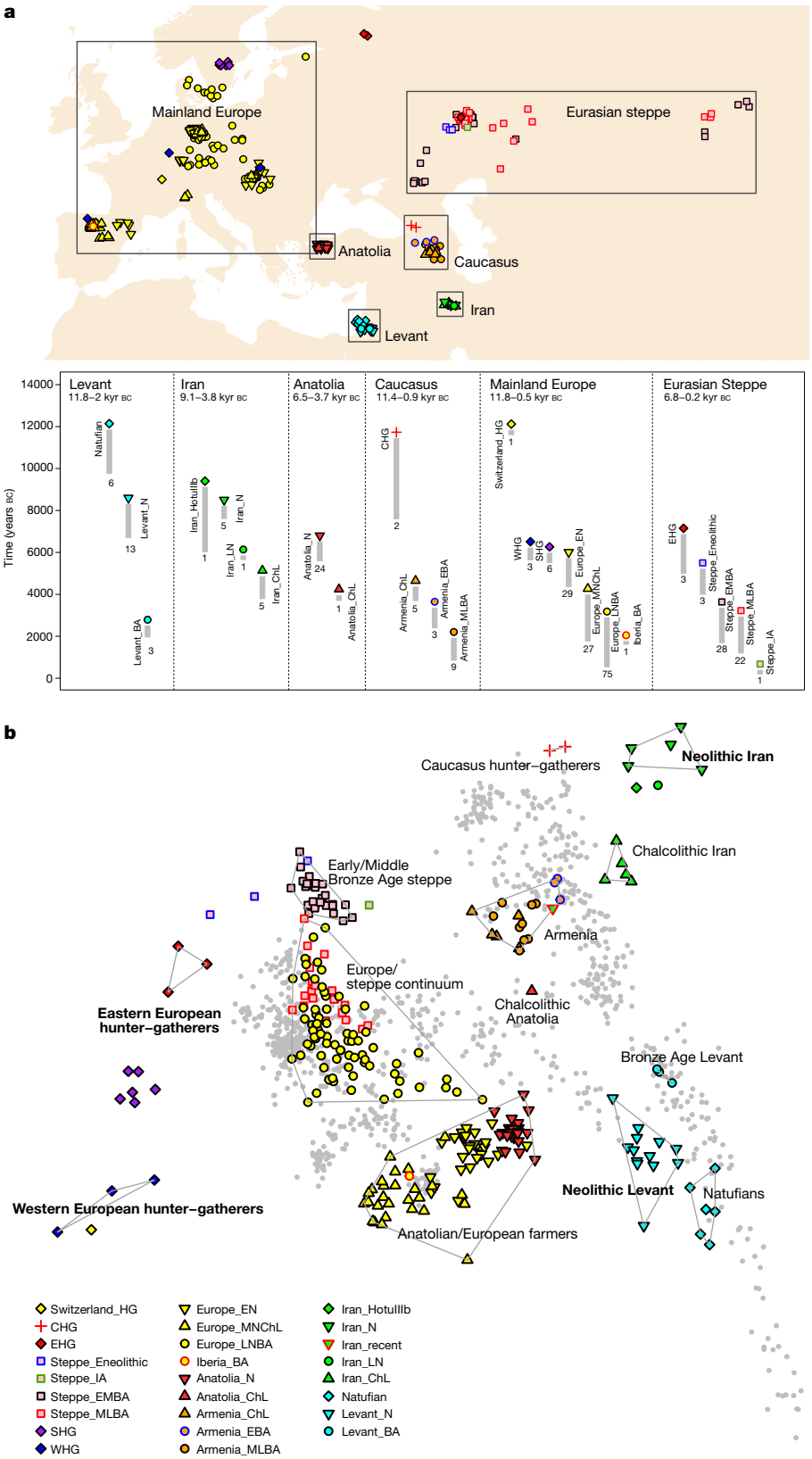


Figure 1 | Genetic structure of ancient West Eurasia. a, Sampling locations and times in six regions. Sample sizes for each population are given below each bar. E, Early; M, Middle; L, Late; HG, hunter–gatherer; N, Neolithic; ChL, Chalcolithic; BA, Bronze Age; IA, Iron Age. **b**, Principal components analysis of 991 present-day West Eurasians (grey points) with 278 projected ancient samples (excluding the Upper Palaeolithic Ust'-Ishim, Kostenki14, and MA1). To avoid visual clutter, population labels of present-day individuals are shown in Extended Data Fig. 1.

quality control for 45 individuals on whom we had a median coverage of 172,819 SNPs. We assembled direct radiocarbon dates on skeletal remains from 26 of these individuals (22 newly generated for this study) (Supplementary Table 1).

The newly reported ancient individuals date to ~12,000–1,400 BC and come from the southern Caucasus (Armenia), northwestern Anatolia (Turkey), Iran, and the southern Levant (Israel and Jordan) (Supplementary Table 1 and Fig. 1a). (One individual had a radiocarbon date that was not in agreement with the date of its archaeological context and was also a genetic outlier.) The samples include Epipalaeolithic Natufian hunter–gatherers from Raqefet Cave in the Levant (~12,000–9,800 BC); a likely Mesolithic individual (HotuIIIb) from Hotu Cave in the Alborz mountains of Iran (probable date of 9,100–8,600 BC); pre-pottery Neolithic farmers from 'Ain Ghazal and

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

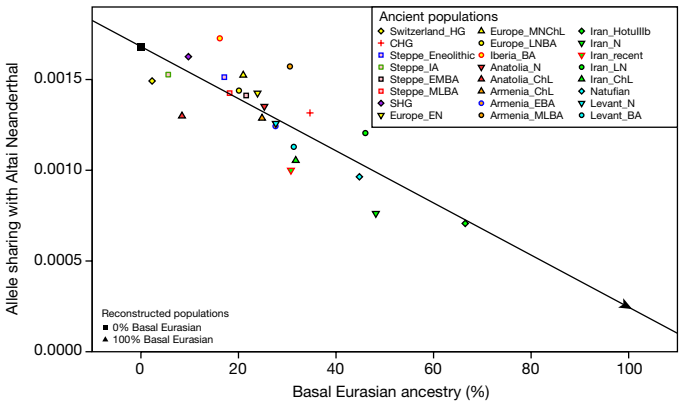


Figure 2 | Basal Eurasian ancestry explains the reduced Neanderthal admixture in West Eurasians. Basal Eurasian ancestry estimates are negatively correlated to a statistic measuring Neanderthal ancestry f_4(*Test*, Mbuti; Altai, Denisovan).

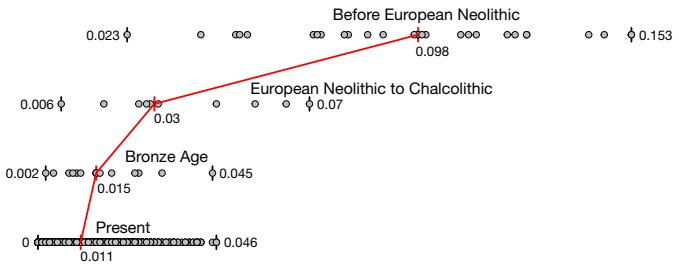


Figure 3 | Genetic differentiation and its marked decrease over time in West Eurasia. Pairwise F_{ST} distribution among populations belonging to four successive time slices in West Eurasia; the median (red) and range of F_{ST} is shown.

Motza in the southern Levant (~8,300–6,700 BC); and early farmers from Ganj Dareh in the Zagros mountains of western Iran (~8,200–7,600 BC). The samples also include later Neolithic, Chalcolithic (~4,800–3,700 BC), and Bronze Age (~3,350–1,400 BC) individuals (Supplementary Information, section 1). We combined our data with previously published ancient data[7–15] to form a dataset of 281 ancient individuals. We then further merged these data with 2,583 present-day people genotyped on the Affymetrix Human Origins array[13,16] (238 newly generated) (Supplementary Table 2 and Supplementary Information, section 2). We grouped the ancient individuals on the basis of archaeological culture and chronology (Fig. 1a and Supplementary Table 1). We refined the grouping on the basis of patterns evident in Principal Components Analysis (PCA)[17] (Fig. 1b and Extended Data Fig. 1), ADMIXTURE model-based clustering[18] (Extended Data Fig. 2a), and 'outgroup' f_3-analysis (Extended Data Fig. 3). We used f_4-statistics to identify outlier individuals and to cluster phylogenetically indistinguishable groups into 'Analysis Labels' (Supplementary Information, section 3).

We analysed these data to address six questions. (1) Previous work has shown that the first European farmers harboured ancestry from a Basal Eurasian lineage that diverged from the ancestors of north Eurasian hunter–gatherers and East Asians before they separated from each other[13]. What was the distribution of Basal Eurasian ancestry in the ancient Near East? (2) Were the first farmers of the Near East part of a single homogeneous population, or were they regionally differentiated? (3) Was there continuity between late pre-agricultural hunter–gatherers and early farming populations, or were the hunter–gatherers largely displaced by a single expansive population, as in early Neolithic Europe?[8] (4) What is the genetic contribution of these early Near Eastern farmers to later populations of the Near East? (5) What is the genetic contribution of the early Near Eastern farmers to later populations of mainland Europe, the Eurasian steppe, and to populations outside West Eurasia? (6) Do our data provide broader insights about population transformations in West Eurasia?

Basal Eurasian and Neanderthal ancestry

The 'Basal Eurasians' are a lineage hypothesized[13] to have split off before the differentiation of all other Eurasian lineages, including eastern non-African populations such as the Han Chinese, and even the early diverged lineage represented by the genome sequence of the ~45,000-year-old Upper Palaeolithic Siberian from Ust'-Ishim[11]. To test for Basal Eurasian ancestry, we computed the statistic f_4(*Test*, Han; Ust'-Ishim, Chimp) (Supplementary Information, section 4), which measures the excess of allele sharing of Ust'-Ishim with a variety of *Test* populations compared to Han as a baseline. This statistic is significantly negative ($Z < -3.7$) for all ancient Near Easterners as well as Neolithic and later Europeans, consistent with them having ancestry from a deeply divergent Eurasian lineage that separated from the ancestors of most Eurasians before the separation of Han and Ust'-Ishim. We used *qpAdm* (ref. 7) to estimate Basal Eurasian ancestry in each *Test* population. We obtained the highest estimates in the earliest populations from both Iran (66 ± 13% in the likely Mesolithic sample, 48 ± 6% in Neolithic samples), and the Levant (44 ± 8% in Epipalaeolithic Natufians) (Fig. 2), showing that Basal Eurasian ancestry was widespread across the ancient Near East.

West Eurasians harbour significantly less Neanderthal ancestry than East Asians[19–21], which could be explained if West Eurasians (but not East Asians) have partial ancestry from a source that diluted their Neanderthal inheritance[20]. Supporting this theory, we observe a negative correlation between Basal Eurasian ancestry and the rate of shared alleles with Neanderthals[19] (Supplementary Information, section 5 and Fig. 2). By extrapolation, we infer that the Basal Eurasian population had lower Neanderthal ancestry than non-Basal Eurasian populations and possibly none (95% confidence interval truncated at zero of 0–60%; Fig. 2; Methods). The finding of little if any Neanderthal ancestry in Basal Eurasians could be explained if the Neanderthal admixture into modern humans ~50,000–60,000 years ago[11] largely occurred after the splitting of the Basal Eurasians from other non-Africans.

It is striking that the highest estimates of Basal Eurasian ancestry are from the Near East, given the hypothesis that it was there that most admixture between Neanderthals and modern humans occurred[19,22]. This could be explained if Basal Eurasians thoroughly admixed into the Near East before the time of the samples we analysed but after the Neanderthal admixture. Alternatively, the ancestors of Basal Eurasians may have always lived in the Near East, but the lineage of which they were a part did not participate in the Neanderthal admixture.

A population without Neanderthal admixture, basal to other Eurasians, may have plausibly lived in Africa. Craniometric analyses have suggested an affinity between the Natufians and populations of north or sub-Saharan Africa[23,24], a result that finds some support from Y chromosome analysis showing that the Natufians and successor Levantine Neolithic populations carried haplogroup E, likely to be of ultimately African origin, which has not been detected in other ancient males from West Eurasia[7,8] (Supplementary Information, section 6). However, no affinity of Natufians to sub-Saharan Africans is evident in our genome-wide analysis, as present-day sub-Saharan Africans do not share more alleles with Natufians than with other ancient Eurasians (Extended Data Table 1). (We could not test for a link to present-day North Africans, who owe most of their ancestry to back-migration from Eurasia[25,26].) The idea of Natufians as a vector for the movement of Basal Eurasian ancestry into the Near East is also not supported by our data, as the Basal Eurasian ancestry in the Natufians (44 ± 8%) is consistent with stemming from the same population as that in the Neolithic and Mesolithic populations of Iran, and is not greater than in those populations (Supplementary Information, section 4). Further insight into the origins and legacy of the Natufians could come from comparison to Natufians from additional sites, and to ancient DNA from North Africa.

Extreme differentiation in the ancient Near East

PCA on present-day West Eurasian populations (Methods and Extended Data Fig. 1), on which we projected the ancient individuals

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

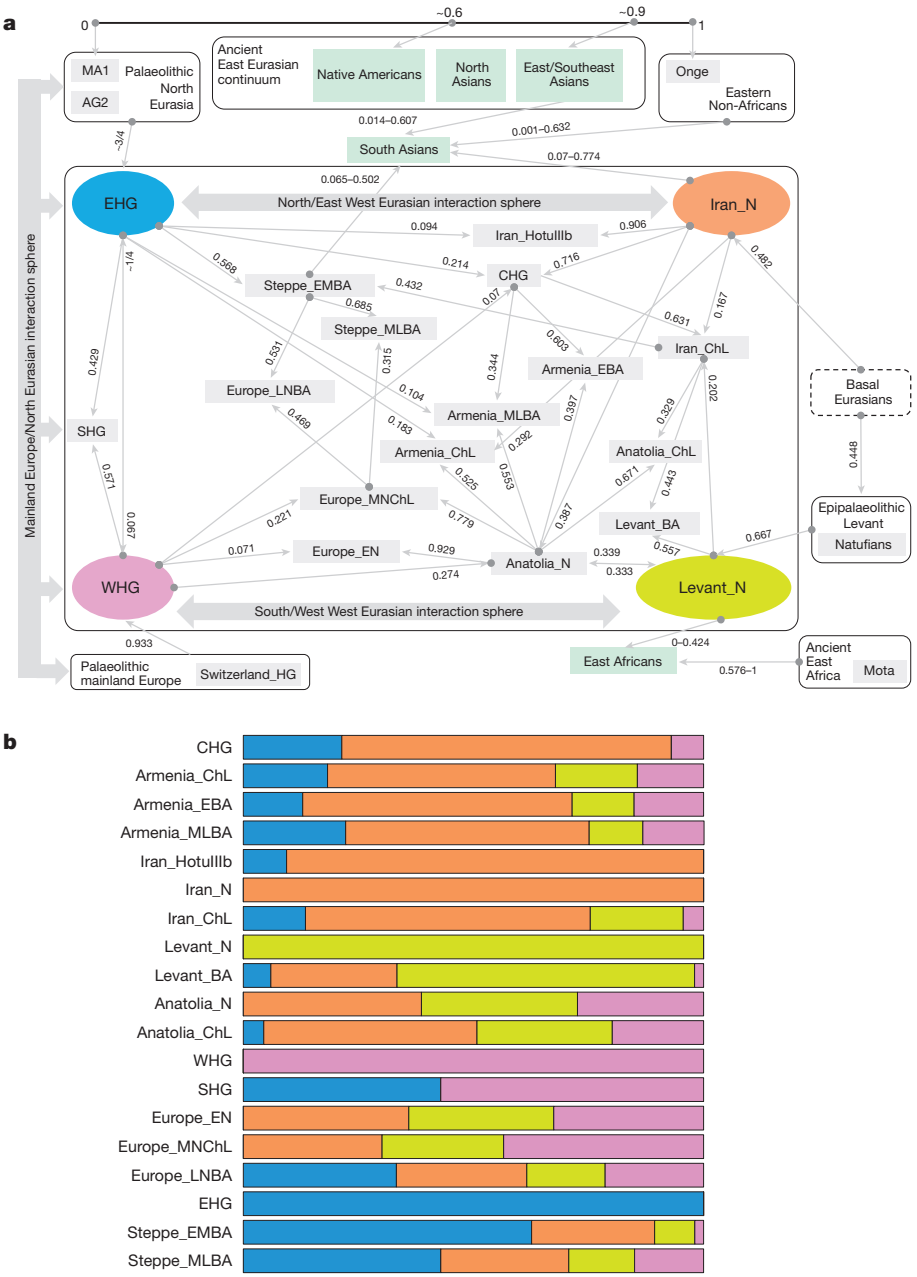


Figure 4 | Modelling ancient West Eurasians, East Africans, East Eurasians and South Asians. a, All the ancient populations can be modelled as mixtures of two or three other populations and up to four proximate sources (marked in colour). Mixture proportions inferred by *qpAdm* are indicated by the incoming arrows to each population. Green rectangles represent sets of more than one population. Multiple admixture solutions are consistent with the data for some populations, and while only one solution is shown here, Supplementary Information, section 7 presents the others. **b**, A flat representation of the graph showing mixture proportions from the four proximate sources.

(Fig. 1b), replicates previous findings of a Europe–Near East contrast along the horizontal principal component 1 (PC1) and parallel clines (PC2) in both Europe and the Near East[7,8,13] (Extended Data Fig. 1). Ancient samples from the Levant clustered at one end of the Near Eastern cline, and ancient samples from Iran at the other. The two Caucasus hunter–gatherers (CHG)[9] are less extreme along PC1 than the Mesolithic and Neolithic individuals from Iran, while individuals from Chalcolithic Anatolia, Iran, Armenia, and Bronze Age Armenia occupy intermediate positions. Qualitatively, the PCA has the appearance of a quadrangle whose four corners are some of the oldest samples: bottom-left, Western hunter–gatherers (WHG); top-left, Eastern hunter–gatherers (EHG); bottom-right, Neolithic Levant and Natufians; top-right, Neolithic Iran. This suggests that diverse ancient West Eurasians can be modelled as mixtures of as few as four streams of ancestry related to these populations, which we confirmed using *qpWave* (ref. 7) (Supplementary Information, section 7).

We computed squared allele frequency differentiation between all pairs of ancient West Eurasians[27] (Methods; Fig. 3 and Extended Data Figs 2b and 4), and found that the populations at the four corners of the quadrangle had differentiation of $F_{ST} = 0.08–0.15$, comparable to the value of 0.09–0.13 seen between present-day West Eurasians and East Asians (Han) (Supplementary Table 3). By contrast, by the Bronze Age, genetic differentiation between pairs of West Eurasian populations had reached its present-day low levels (Fig. 3): today, F_{ST} is ≤ 0.025 for 95% of the pairs of West Eurasian populations and ≤ 0.046 for all pairs (Supplementary Table 3). These results point to a demographic process that established high differentiation across West Eurasia and then reduced this differentiation over time.

Continuity between hunter–gatherers and early farmers

Our data document continuity across the transition between hunter–gatherers and farmers, separately in the southern Levant and in the southern Caucasus–Iran highlands. The qualitative evidence for this is that PCA, ADMIXTURE, and outgroup f_3 analysis cluster Levantine hunter–gatherers (Natufians) with Levantine farmers, and Iranian and CHG with Iranian farmers (Fig. 1b and Extended Data Figs 1, 3). We confirm this in the Levant by showing that its early farmers share significantly more alleles with Natufians than with the early farmers of Iran:

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

the statistic f_4(Levant_N, Chimp; Natufian, Iran_N) is significantly positive ($Z = 13.6$). The early farmers of the Caucasus–Iran highlands similarly share significantly more alleles with the hunter–gatherers of this region than with the early farmers from the Levant: the statistic f_4(Iran_N, Chimp; Caucasus or Iran highland hunter–gatherers, Levant_N) is significantly positive ($Z > 6$).

Admixture in the ancient Near East

Almost all ancient and present-day West Eurasians have evidence of significant admixture between two or more ancestral populations, as documented by statistics of the form f_3(*Test*; *Reference*$_1$, *Reference*$_2$) which, if negative, show that a test population's allele frequencies tend to be an intermediate between two reference populations[16] (Extended Data Table 2). To better understand the admixture history beyond these patterns, we used *qpAdm* (ref. 7), which can evaluate whether a particular test population is consistent with being derived from a set of proposed source populations, and if so, infer mixture proportions (Methods). We used this approach to carry out a systematic survey of ancient West Eurasian populations to explore their possible sources of admixture (Fig. 4 and Supplementary Information, section 7).

Among first farmers, those of the Levant trace approximately two-thirds of their ancestry to people related to Natufian hunter–gatherers and about one-third to people related to Anatolian farmers (Supplementary Information, section 7). Western Iranian first farmers cluster with the likely Mesolithic HotuIIIb individual and more remotely with hunter–gatherers from the southern Caucasus (Fig. 1b), and share alleles at an equal rate with Anatolian and Levantine early farmers (Supplementary Information, section 7), highlighting the long-term isolation of western Iran.

During subsequent millennia, the early farmer populations of the Near East expanded in all directions and mixed, as we can model populations of the Chalcolithic and subsequent Bronze Age only as having ancestry from two or more sources. The Chalcolithic people of western Iran can be modelled as a mixture of the Neolithic people of western Iran, the Levant and CHG, consistent with their position in the PCA (Fig. 1b). Admixture from populations related to the Chalcolithic people of western Iran had a wide impact, consistent with contributing around 44% of the ancestry of Levantine Bronze Age populations in the south and about 33% of the ancestry of the Chalcolithic North-West Anatolians in the west. Our analysis shows that the ancient populations of Chalcolithic Iran, Chalcolithic Armenia, Bronze Age Armenia and Chalcolithic Anatolia were all composed of the same ancestral components, albeit in slightly different proportions (Fig. 4b and Supplementary Information, section 7).

Admixture into Europe, East Africa and South Asia

Admixture did not only occur within the Near East but also extended towards Europe. To the north, a population related to people of Chalcolithic Iran contributed about 43% of the ancestry of early Bronze Age populations of the steppe. The spread of Near Eastern ancestry into the Eurasian steppe was previously inferred[7] without access to ancient samples, with a population related to present-day Armenians as a suggested source[7,8]. To the west, the early farmers of mainland Europe were descended from a population related to Neolithic North-Western Anatolians[8]. This is consistent with an Anatolian origin of farming in Europe, but does not reject other sources, as the spatial distribution of the Anatolian/European-like farmer populations is unknown. We can rule out the hypothesis that European farmers stem directly from a population related to the ancient farmers of the southern Levant[28,29], however, because European farmers share more alleles with Anatolian Neolithic farmers than with Levantine farmers, as attested by the positive statistic f_4(Europe_EN, Chimp; Anatolia_N, Levant_N) ($Z = 15$).

Migration from the Near East also occurred towards the southwest into East African populations, which experienced West Eurasian admixture around 1,000 BC[30,31]. Previously, the West Eurasian population known to be the best proxy for this ancestry was present-day Sardinians[31], who resemble Neolithic Europeans genetically[13,32]. However, our analysis shows that East African ancestry is significantly better modelled by Levantine early farmers than by Anatolian or early European farmers, implying that the spread of this ancestry to East Africa was not from the same group that spread Near Eastern ancestry into Europe (Extended Data Fig. 5 and Supplementary Information, section 8).

In South Asia, our dataset provides insight into the sources of Ancestral North Indians (ANI), a West Eurasian-related population that no longer exists in unmixed form but contributes a variable amount of the ancestry of South Asians[33,34] (Supplementary Information, section 9 and Extended Data Fig. 5). We show that it is impossible to model the ANI as being derived from any single ancient population in our dataset. However, it can be modelled as a mix of ancestries related to both early farmers of western Iran and people of the Bronze Age Eurasian steppe; all sampled South Asian groups are inferred to have significant amounts of both ancestral types. The demographic impact of steppe-related populations on South Asia was substantial, as the Mala, a south Indian population with minimal ANI along the 'Indian Cline' of such ancestry[33,34], is inferred to have around 18% steppe-related ancestry, while the Kalash of Pakistan are inferred to have about 50%, similar to present-day northern Europeans[7].

Population transformations in West Eurasia and beyond

We were concerned that our conclusions might be biased by the particular populations we happened to sample, and that we would have obtained qualitatively different conclusions without data from some key populations. We tested our conclusions by plotting the inferred position of admixed populations in PCA against a weighted combination of their inferred source populations and obtained qualitatively consistent results (Extended Data Fig. 6).

To further assess the robustness of our inferences, we developed a method to infer the existence and genetic affinities of ancient populations from unobserved 'ghost' populations (Supplementary Information, section 10 and Extended Data Fig. 7). This method takes advantage of the insight that if an unsampled ghost population admixes with differentiated 'substratum' populations, it is possible to extrapolate its identity by intersecting clines of populations with variable proportions of ghost and substratum ancestry. Applying this approach while withholding major populations, we validated some of our key inferences, successfully inferring mixture proportions consistent with those obtained when the populations were included in the analysis. Application of this method highlights the impact of Ancient North Eurasian (ANE) ancestry related to the ~22,000 BC Mal'ta 1 and ~15,000 BC Afontova Gora 2 (ref. 15) on populations living in Europe, the Americas and Eastern Eurasia. Eastern Eurasians can be modelled as arrayed along a cline with different proportions of ANE ancestry (Supplementary Information, section 11 and Extended Data Fig. 8), ranging from about 40% ANE in Native Americans, matching previous findings[13,15], to no less than around 5–10% ANE in diverse East Asian groups including Han Chinese (Extended Data Figs 5, 7f). We also document a cline of ANE ancestry across the East–West extent of Eurasia. Eastern hunter-gatherers (EHG) derive about three-quarters of their ancestry from the ANE (Supplementary Information, section 11); Scandinavian hunter–gatherers[7,8,13] (SHG) are a mix of EHG and WHG; and WHG are a mix of EHG and populations related to the Upper Palaeolithic Bichon from Switzerland (Supplementary Information, section 7). Northwest Anatolians—with ancestry from a population related to European hunter–gatherers (Supplementary Information, section 7)—are better modelled if this ancestry is taken as more extreme than Bichon (Supplementary Information, section 10).

The population structure of the ancient Near East was not independent of that of Europe (Supplementary Information, section 4), as evidenced by the highly significant ($Z = -8.9$) statistic f_4(Iran_N, Natufian;WHG, EHG) which suggests gene flow in 'northeastern' (Neolithic Iran/EHG) and 'southwestern' (Levant/WHG) interaction spheres (Fig. 4d). This interdependence of the ancestry of Europe and

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

the Near East may have been mediated by unsampled geographically intermediate populations[35] that contributed ancestry to both regions.

Conclusions

By analysing genome-wide ancient DNA data from ancient individuals from the Levant, Anatolia, the southern Caucasus and Iran, we have provided a first glimpse into the demographic structure of the human populations that transitioned to farming. We reject the hypothesis that the spread of agriculture in the Near East was achieved by the dispersal of a single farming population displacing the hunter–gatherers they encountered. Instead, the spread of ideas and farming technology moved faster than the spread of people, as we can determine from the fact that the population structure of the Near East was maintained throughout the transition to agriculture. A priority for future ancient DNA studies should be to obtain data from older periods, which would reveal the deeper origins of the population structure in the Near East. It will also be important to obtain data from the ancient civilizations of the Near East to bridge the gap between the region's prehistoric inhabitants and those of the present.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 8 March; accepted 18 July 2016.
Published online 25 July 2016.

1. Barker, G. & Goucher, C. *The Cambridge World History Volume II: A World with agriculture, 12,000 BCE–500 CE* (Cambridge Univ. Press, 2015).
2. Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. *The History and Geography of Human Genes.* (Princeton Univ. Press, 1994).
3. Gamba, C. *et al.* Genome flux and stasis in a five millennium transect of European prehistory. *Nat. Commun.* **5,** 5257 (2014).
4. Pinhasi, R. *et al.* Optimal ancient DNA yields from the inner ear part of the human petrous bone. *PLoS One* **10,** e0129102 (2015).
5. Fu, Q. *et al.* DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl Acad. Sci. USA* **110,** 2223–2227 (2013).
6. Fu, Q. *et al.* An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524,** 216–219 (2015).
7. Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522,** 207–211 (2015).
8. Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528,** 499–503 (2015).
9. Jones, E. R. *et al.* Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.* **6,** 8912 (2015).
10. Allentoft, M. E. *et al.* Population genomics of Bronze Age Eurasia. *Nature* **522,** 167–172 (2015).
11. Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514,** 445–449 (2014).
12. Günther, T. *et al.* Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. *Proc. Nat. Acad Sci. USA* (2015).
13. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513,** 409–413 (2014).
14. Olalde, I. *et al.* A common genetic origin for early farmers from Mediterranean Cardial and Central European LBK cultures. *Mol. Biol. Evol.* **32,** 3132–3142 (2015).
15. Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505,** 87–91 (2014).
16. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192,** 1065–1093 (2012).
17. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2,** e190 (2006).
18. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19,** 1655–1664 (2009).
19. Prufer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505,** 43–49 (2014).
20. Meyer, M. *et al.* A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338,** 222–226 (2012).
21. Wall, J. D. *et al.* Higher levels of neanderthal ancestry in East Asians than in Europeans. *Genetics* **194,** 199–209 (2013).
22. Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328,** 710–722 (2010).
23. Brace, C. L. *et al.* The questionable contribution of the Neolithic and the Bronze Age to European craniofacial form. *Proc. Natl Acad. Sci. USA* **103,** 242–247 (2006).
24. Ferembach, D. Squelettes du Natoufien d'Israel., etude anthropologique. *Anthropologie* **65,** 46–66 (1961).
25. Fadhlaoui-Zid, K. *et al.* Genome-wide and paternal diversity reveal a recent origin of human populations in North Africa. *PLoS One* **8,** e80293 (2013).
26. Henn, B. M. *et al.* Genomic ancestry of North Africans supports back-to-Africa migrations. *PLoS Genet.* **8,** e1002397 (2012).
27. Bhatia, G., Patterson, N., Sankararaman, S. & Price, A. L. Estimating and interpreting FST: the impact of rare variants. *Genome Res.* **23,** 1514–1521 (2013).
28. Fernández, E. *et al.* Ancient DNA analysis of 8000 B.C. near eastern farmers supports an early neolithic pioneer maritime colonization of Mainland Europe through Cyprus and the Aegean Islands. *PLoS Genet.* **10,** e1004401 (2014).
29. Ammerman, A. J., Pinhasi, R. & Banffy, E. Comment on Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. *Science* **312,** 1875; author reply 1875 (2006).
30. Pagani, L. *et al.* Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool. *Am. J. Hum. Genet.* **91,** 83–96 (2012).
31. Pickrell, J. K. *et al.* Ancient west Eurasian ancestry in southern and eastern Africa. *Proc. Natl Acad. Sci. USA* **111,** 2632–2637 (2014).
32. Keller, A. *et al.* New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. *Nat Commun* **3,** 698 (2012).
33. Moorjani, P. *et al.* Genetic evidence for recent population mixture in India. *Am. J. Hum. Genet.* **93,** 422–438 (2013).
34. Reich, D., Thangaraj, K., Patterson, N., Price, A. L. & Singh, L. Reconstructing Indian population history. *Nature* **461,** 489–494 (2009).
35. Fu, Q. *et al.* The genetic history of Ice Age Europe. *Nature* **534,** 200–205 (2016).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the 238 human subjects who donated samples for genome-wide analysis, and D. Labuda and P. Zalloua for sharing samples from Poland and Lebanon. The Fig. 1a map was plotted in R using the worldHiRes map of the 'mapdata' package (using public domain data from the CIA World Data Bank II). We thank O. Bar-Yosef, M. Bonogofsky, I. Hershkowitz, M. Lipson, I. Mathieson, H. May, R. Meadow, I. Olalde, S. Paabo, P. Skoglund, and N. Nakatsuka for comments and critiques, and D. Bradley, M. Dallakyan, S. Esoyan, M. Ferry and M. Michel, and A. Yesayan, for contributions to bone preparation and ancient DNA work. D.F. and M.N. were supported by Irish Research Council grants GOIPG/2013/36 and GOIPD/2013/1, respectively. S.C. was funded by the Irish Research Council for Humanities and Social Sciences (IRCHSS) ERC Support Programme. Q.F. was funded by the Bureau of International Cooperation of the Chinese Academy of Sciences, the National Natural Science Foundation of China (L1524016) and the Chinese Academy of Sciences Discipline Development Strategy Project (2015-DX-C-03). The Scottish diversity data was funded by the Chief Scientist Office of the Scottish Government Health Directorates (CZD/16/6), the Scottish Funding Council (HR03006), and a project grant from the Scottish Executive Health Department, Chief Scientist Office (CZB/4/285). M.S., A.Tön., M.B. and P.K. were supported by the German Research Foundation (CRC 1052; B01, B03, C01). M.S.-P. was funded by a Wenner-Gren Foundation Dissertation Fieldwork grant (9005), and by the National Science Foundation DDRIG (BCS-1455744). P.K. was funded by the Federal Ministry of Education and Research, Germany (FKZ: 01EO1501). J.F.W. acknowledge the MRC 'QTL in Health and Disease' programme grant. The Romanian diversity data was supported by the EC Commission, Directorate General XII (Supplementary Agreement ERBCIPDCT 940038 to the Contract ERBCHRXCT 920032, coordinated by A. Piazza, Turin, Italy). M.R. received support from the Leverhulme Trust's Doctoral Scholarship programme. O.S. and A.Tor. were supported by the University of Pavia (MIGRAT-IN-G) and the Italian Ministry of Education, University and Research: Progetti Ricerca Interesse Nazionale 2012. The Raqefet Cave Natufian project was supported by funds from the National Geographic Society (grant 8915-11), the Wenner-Gren Foundation (grant 7481) and the Irene Levi-Sala CARE Foundation, while radiocarbon dating on the samples was funded by the Israel Science Foundation (grant 475/10; E. Boaretto). R.P. was supported by ERC starting grant ADNABIOARC (263441). D.R. was supported by NIH grant GM100233, by NSF HOMINID BCS-1032255, and is a Howard Hughes Medical Institute investigator.

Author Contributions R.P. and D.R. conceived the idea for the study. D.N., G.R., D.C.M., S.C., S.A.R., G.L., F.B., B.Gas., J.M.M., M.G., V.E., A.M., C.M., F.G., N.A.H. and R.P. assembled skeletal material. N.R., D.F., M.N., B.Gam., K.Si., S.C., K.St., E.H., Q.F., G.G.-F., E.R.J., R.P. and D.R. performed or supervised ancient DNA wet laboratory work. L.B, M.B., A.C., G.C., D.C., P.F., E.G., S.M.K., P.K., J.K., D.M., M.M., D.A.M., S.O., M.B.R., O.S., M.S.-P., G.S., M.S., A.Tön., A.Tor., J.F.W., L.Y. and D.R. assembled present-day samples for genotyping. I.L, N.P. and D.R. developed methods for data analysis. I.L., S.M., Q.F., N.P. and D.R. analysed data. I.L., R.P. and D.R. wrote the manuscript and supplements.

Author Information The aligned sequences are available through the European Nucleotide Archive under accession number PRJEB14455. Fully public subsets of the analysis datasets are at http://genetics.med.harvard.edu/reichlab/Reich_Lab/Datasets.html. The complete dataset (including present-day humans for which the informed consent is not consistent with public posting of data) is available to researchers who send a signed letter to D.R. indicating that they will abide by specified usage conditions (Supplementary Information, section 2). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to I.L. (lazaridis@genetics.med.harvard.edu) or R.P. (ron.pinhasi@ucd.ie) or D.R. (reich@genetics.med.harvard.edu).

Reviewer Information *Nature* thanks O. Bar-Yosef, G. Coop and the other anonymous reviewer(s) for their contribution to the peer review of this work.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Ancient DNA data. In a dedicated ancient DNA laboratory at University College Dublin, we prepared powder from 132 ancient Near Eastern samples, either by dissecting the inner ear region of the petrous bone using a sandblaster (Renfert), or by drilling using a Dremel tool and single-use drill bits and selecting the best preserved bone fragments based on anatomical criteria. These fragments were then powdered using a mixer mill (Retsch Mixer Mill 400)[4].

We performed all subsequent processing steps in a dedicated ancient DNA laboratory at Harvard Medical School, where we extracted DNA from the powder (usually 75 mg, range 14–81 mg) using an optimized ancient DNA extraction protocol[36], but replaced the assembly of Qiagen MinElute columns and extension reservoirs from Zymo Research with a High Pure Extender Assembly from the High Pure Viral Nucleic Acid Large Volume Kit (Roche Applied Science). We built a total of 170 barcoded double-stranded Illumina sequencing libraries for these samples[37], of which we treated 167 with uracil-DNA glycosylase (UDG) to remove the characteristic C-to-T errors of ancient DNA[38]. The UDG treatment strategy is (by–design) inefficient at removing terminal uracils, allowing the mismatch rate to the human genome at the terminal nucleotide to be used for authentication[37]. We updated this library preparation protocol in two ways compared to the original publication: first, we used 16U Bst2.0 Polymerase, Large Fragment (NEB) and 1× Isothermal amplification buffer (NEB) in a final volume of 25 μl fill-in reaction, and second, we used the entire inactivated 25 μl fill-in reaction in a total volume of 100 μl PCR mix with 1 μM of each primer[39]. We included extraction negative controls (where no sample powder was used) and library negative controls (where extract was supplemented by water) in every batch of samples processed and carried them through the entire wet laboratory processing to test for reagent contamination.

We screened the libraries by hybridizing them in solution to a set of oligonucleotide probes tiling the mitochondrial genome[40], using the protocol described previously[7]. We sequenced the enriched libraries using an Illumina NextSeq 500 instrument using 2× 76 bp reads, trimmed identifying sequences (seven base pair molecular barcodes at either end) and any trailing adapters, merged read pairs that overlapped by at least 15 base pairs, and mapped the merged sequences to the RSRS mitochondrial DNA reference genome[41], using the Burrows Wheeler Aligner[42] (*bwa*) and the command *samse* (v0.6.1).

We enriched promising libraries for a targeted set of ∼1.2 million SNPs[8] as in ref. 5, and adjusted the blocking oligonucleotide and primers to be appropriate for our libraries. The specific probe sequences are given in supplementary data 2 of ref. 7. and supplementary data 1 of ref. 6. We sequenced the libraries on an Illumina NextSeq 500 using 2× 76 bp reads. We trimmed identifying sequences (molecular barcodes) and any trailing adapters, merged pairs that overlapped by at least 15 base pairs (allowing up to one mismatch), and mapped the merged sequences to *hg19* using the single-ended aligner *samse* in bwa (v0.6.1). We removed duplicated sequences by identifying sets of sequences with the same orientation and start and end positions after alignment to *hg19*; we picked the highest quality sequence to represent each set. For each sample, we represented each SNP position by a randomly chosen sequence, restricting to sequences with a minimum mapping quality (MAPQ ≥ 10), sites with a minimum sequencing quality (≥20), and removing two bases at the ends of reads. We sequenced the enriched products up to the point that we estimated that generating a hundred new sequences was expected to add data on less than about one new SNP[8].

Testing for contamination and quality control. For each ancient DNA library, we evaluated authenticity in several ways. First, we estimated the rate of matching to the consensus sequence for mitochondrial genomes sequenced to a coverage of at least tenfold from the initial screening data. Of the 76 libraries that contributed to our dataset (coming from 45 samples), 70 had an estimated rate of sequencing matching to the consensus of >95% according to contamMix[5] (the remaining libraries had estimated match rates of 75–92%, but gave no sign of being outliers in principal component analysis or X-chromosome contamination analysis so we retained them for analysis) (Supplementary Table 1). We quantified the rate of C-to-T substitution in the final nucleotide of the sequences analysed, relative to the human reference genome sequence, and found that all the libraries analysed had rates of at least 3% (ref. 37), consistent with genuine ancient DNA. For the nuclear data from males, we used the ANGSD software[43] to obtain a conservative X-chromosome estimate of contamination. We determined that all libraries that passed our quality control and for which we had sufficient X-chromosome data to make an assessment, had contamination rates of 0–1.5%. Finally, we merged data for samples for which we had multiple libraries to produce an analysis dataset.

Affymetrix Human Origins genotyping data. We genotyped 238 present-day individuals from 17 diverse West Eurasian populations on the Affymetrix Human Origins array[16], and applied quality control analyses as previously described[13] (Supplementary Table 2). We merged the newly generated data with data from 2,345 individuals previously genotyped on the same array[13]. All individuals that were genotyped provided individual informed consent consistent with studies of population history, following protocols approved by the ethical review committees of the institutions of the researchers who collected the samples. The collection and analysis of genome-wide data on anonymized samples at Harvard Medical School for the purpose of studying population history was approved by the Harvard Human Research Protection Program, protocol 11681, re-reviewed on 12 July 2016. Anonymized aliquots of DNA from all individuals were sent to the core facility of the Center for Applied Genomics at the Children's Hospital of Philadelphia for genotyping and data processing. For 127 of the individuals with newly reported data, the informed consent was consistent with public distribution of data, and the data can be downloaded at http://genetics.med.harvard.edu/reich/Reich_Lab/Datasets.html. To access data for the remaining 111 newly reported samples, researchers should send a signed letter to D.R. containing the following text: "(a) I will not distribute the data outside my collaboration; (b) I will not post the data publicly; (c) I will make no attempt to connect the genetic data to personal identifiers for the samples; (d) I will use the data only for studies of population history; (e) I will not use the data for any selection studies; (f) I will not use the data for medical or disease-related analyses; (g) I will not use the data for commercial purposes." Supplementary Table 2 specifies which samples are consistent with which type of data distribution.

Datasets. We carried out population genetic analysis on two datasets: (i) *HO* includes 2,583 present-day humans genotyped on the Human Origins array[13,16] including 238 newly reported, (Supplementary Table 2; Supplementary Information, section 2), and 281 ancient individuals on a total of 592,146 autosomal SNPs. (ii) *HOIll* includes the 281 ancient individuals on a total of 1,055,186 autosomal SNPs, including those present in both the Human Origins and Illumina genotyping platforms, but excluding SNPs on the sex chromosomes or additional SNPs of the 1,240k capture array that were included because of their potential functional importance[8]. We used *HO* for analyses that involve both ancient and present-day individuals, and *HOIll* for analysis on ancient individuals alone. We also used 235 individuals from Pagani *et al.*[30] genotyped at 418,700 autosomal SNPs to study admixture in East Africans (Supplementary Information, section 8). Ancient individuals are represented in 'pseudo-haploid' form by randomly choosing one allele for each position of the array.

Principal components analysis. We carried out principal components analysis in the *smartpca* program of EIGENSOFT[17], using default parameters and the lsqproject: YES[13] and numoutlieriter: 0 options. We carried out PCA on the *HO* dataset for 991 present-day West Eurasians (Extended Data Fig. 1), and projected the 278 ancient individuals (Fig. 1b).

ADMIXTURE analysis. We carried out ADMIXTURE analysis[18] of the *HO* dataset after pruning for linkage disequilibrium in PLINK[44,45] with parameters indep-pairwise 200 25 0.4, which retained 296,309 SNPs. We performed analysis in 20 replicates with different random seeds, and retained the highest likelihood replicate for each value of *K*. We show the $K = 11$ results for the 281 ancient samples in Extended Data Fig. 2a (this is the lowest *K* for which components maximized in European hunter–gatherers, ancient Levant, and ancient Iran appear).

***f*-statistics.** We carried out analysis of f_3-statistics, f_4-ratio, and f_4-statistics statistics using the ADMIXTOOLS[16] programs *qp3Pop*, *qpF4ratio* with default parameters, and *qpDstat* with f4mode: YES, and computed standard errors with a block jack-knife[46]. For computing f_3-statistics with an ancient population as a target, we set the inbreed: YES parameter. We computed *f*-statistics on the *HOIll* dataset when no present-day humans were involved and on the *HO* dataset when they were. We computed the statistic f_4(*Test*, Mbuti; Altai, Denisovan) in Fig. 2 on the *HOIll* dataset after merging with whole genome data on 3 Mbuti individuals from Panel C of the Simons Genome Diversity Project[47]. We computed the dendrogram of Extended Data Fig. 3 showing hierarchical clustering of populations with outgroup f_3-statistics using the open source *heatmap.2* function of the *gplots* package in *R*.

Negative correlation of Basal Eurasian ancestry with Neanderthal ancestry. We used the *lm* function of R to fit a linear regression of the rate of allele sharing of a *Test* population with the Altai Neanderthal as measured by f_4(Test, Mbuti; Altai, Denisovan) as the dependent variable, and the proportion of Basal Eurasian ancestry (Supplementary Information, section 4) as the predictor variable. Extrapolating from the fitted line, we obtain the value of the statistic expected if *Test* is a population of 0% or 100% Basal Eurasian ancestry. We then compute the ratio of the Neanderthal ancestry estimate in Basal Eurasians relative to non-Basal Eurasians as f_4(100% Basal Eurasian, Mbuti; Altai, Denisovan)/ f_4(0% Basal Eurasian, Mbuti; Altai, Denisovan). We use a block jack-knife[46], dropping one of 100 contiguous blocks of the genome at a time, to estimate the value and standard error of this quantity (9 ± 26%). We compute a 95% confidence interval based on the point estimate ± 1.96-times the standard error: −42 to 60%. We truncated to 0–60%

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

on the assumption that Basal Eurasians had no less Neanderthal admixture than Mbuti from sub-Saharan Africa.

Estimation of F_{ST} coefficients. We estimated F_{ST} in *smartpca*[17] with default parameters, inbreed: YES, and fstonly: YES.

Admixture graph modelling. We carried out Admixture Graph modelling with the *qpGraph* software[16] using Mbuti as an outgroup unless otherwise specified.

Testing for the number of streams of ancestry. We used the *qpWave*[33,48] software, described in Supplementary Information, section 10 of ref. 7, to test whether a set of 'Left' populations is consistent with being related via as few as N streams of ancestry to a set of 'Right' populations by studying statistics of the form $X(u, v) = F_4(u_0, u; v_0, v)$ where u_0, v_0 are basis populations chosen from the 'Left' and 'Right' sets and u, v are other populations from these sets. We use a Hotelling's T^2 test[48] to evaluate whether the matrix of size $(L-1)*(R-1)$, where L, R are the sizes of the 'Left' and 'Right' sets has rank m. If this is the case, we can conclude that the 'Left' set is related via at least $N = m+1$ streams of ancestry differently to the 'Right' set. We use the parameter allsnps: YES which computes each f_4-statistic based on the full set of SNPs with coverage among the four populations used in the statistic (without regard to whether the SNPs are covered in the other populations in the 'Left' and 'Right' sets).

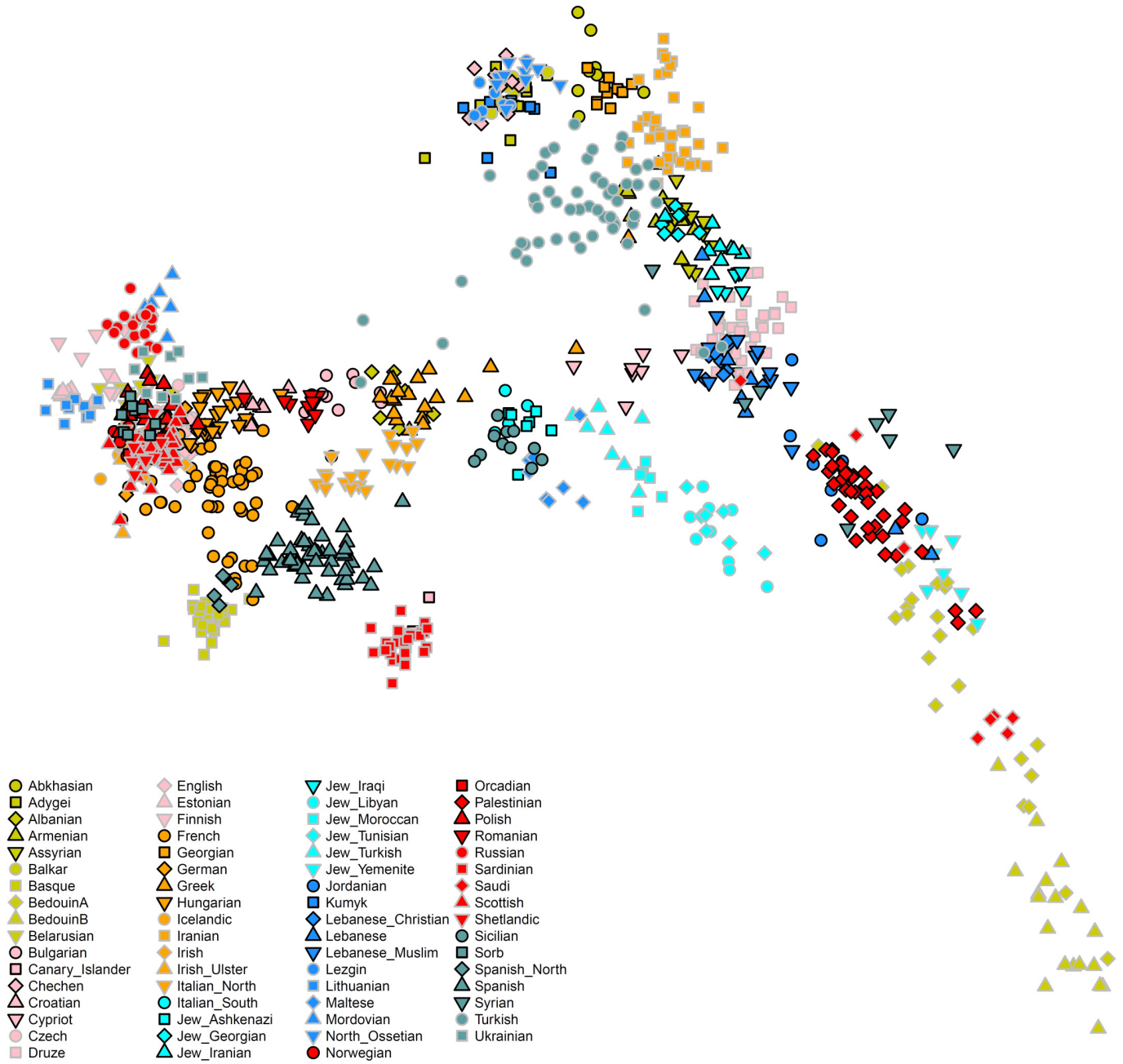
Inferring mixture proportions without an explicit phylogeny. We used the *qpAdm* methodology described in Supplementary Information, section 10 of ref. 7 to estimate the proportions of ancestry in a *Test* population deriving from a mixture of N 'reference' populations by exploiting (but not explicitly modelling) shared genetic drift with a set of 'Outgroup' populations (Supplementary Information, section 7). We set the details: YES parameter, which reports a normally distributed Z-score estimated with a block jack-knife for the difference between the statistics $f_4(u_0, Test; v_0, v)$ and $f_4(u_0, Estimated\ Test; v_0, v)$ where *Estimated Test* is $\sum_{i=1}^{N} \alpha_i f_4(u_0, Ref_i; v_0, v)$, the average of these f_4-statistics weighed by the mixture proportions α_i from the N reference populations. We use the allsnps: YES parameter.

Modelling admixture from ghost populations. We model admixture from a 'ghost' (unobserved) population X in the specific case that X has part of its ancestry from two unobserved ancestral populations p and q. Any population X composed of the same populations p and q resides on a line defined by two observed reference populations r_1 and r_2 composed of the same elements p and q according to a parametric equation $x = r_1 + \lambda(r_2 - r_1)$ with real-valued parameter λ. We define and solve the optimization problem of fitting λ and obtain mixture proportions (Supplementary Information, section 10).

Code availability. Code implementing the newly developed method for modelling admixture from ghost populations is available on request from I.L.

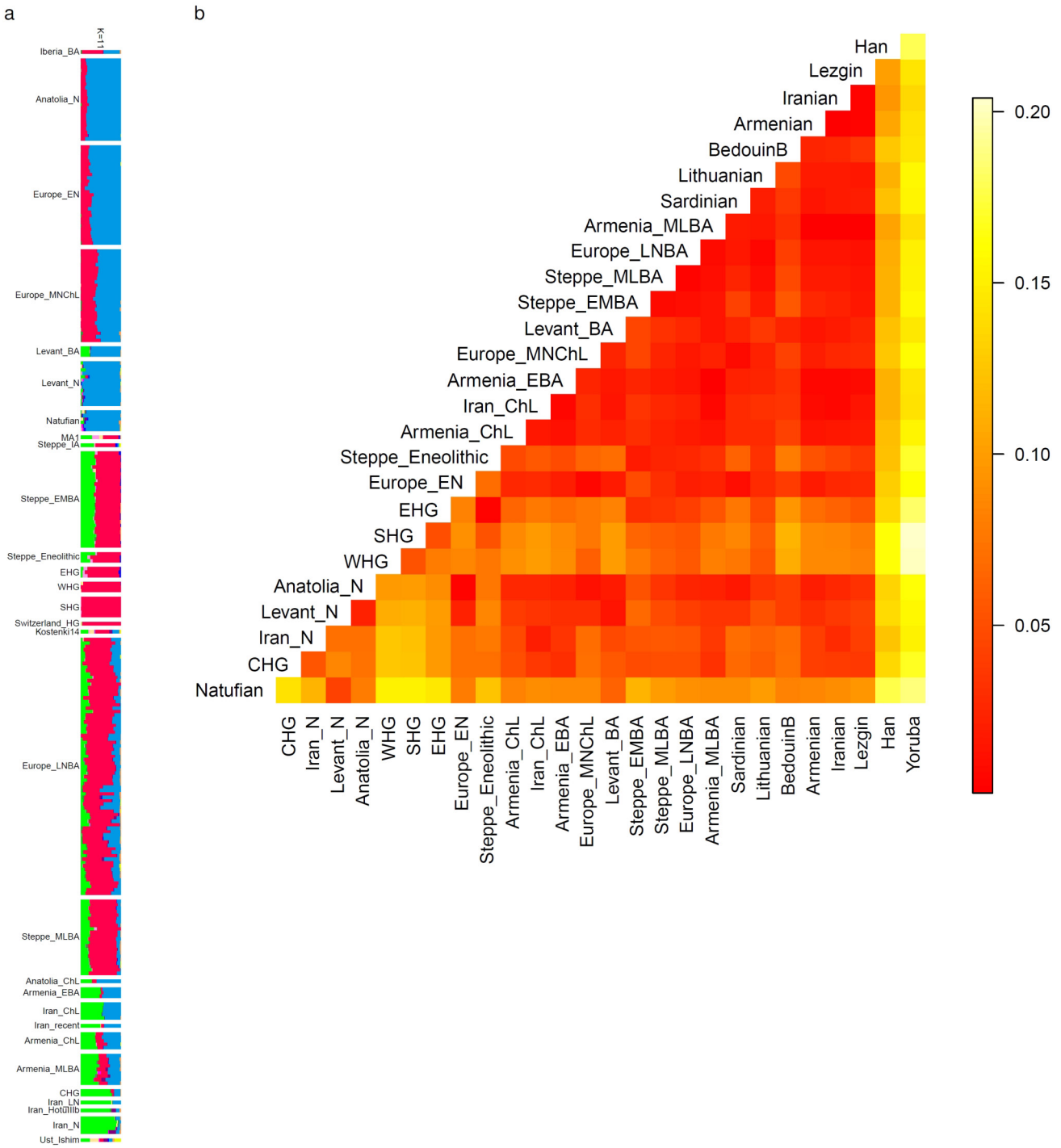
36. Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl Acad. Sci. USA* **110,** 15758–15763 (2013).
37. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Phil. Trans. R. Soc. Lond. B* **370,** 20130624 (2015).
38. Briggs, A. W. *et al.* Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res.* **38,** e87 (2010).
39. Korlević, P. *et al.* Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques* **59,** 87–93 (2015).
40. Meyer, M. *et al.* A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature* **505,** 403–406 (2014).
41. Behar, D. M. *et al.* A "Copernican" reassessment of the human mitochondrial DNA tree from its root. *Am. J. Hum. Genet.* **90,** 675–684 (2012).
42. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
43. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics* **15,** 356 (2014).
44. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81,** 559–575 (2007).
45. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4,** 7 (2015).
46. Busing, F. T. A., Meijer, E. & Leeden, R. Delete-m Jackknife for Unequal m. *Stat. Comput.* **9,** 3–8 (1999).
47. Sudmant, P. H. *et al.* Global diversity, population stratification, and selection of human copy-number variation. *Science* **349,** aab3761 (2015).
48. Reich, D. *et al.* Reconstructing Native American population history. *Nature* **488,** 370–374 (2012).
49. Gallego Llorente, M. *et al.* Ancient Ethiopian genome reveals extensive Eurasian admixture in Eastern Africa. *Science* **350,** 820–822 (2015).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



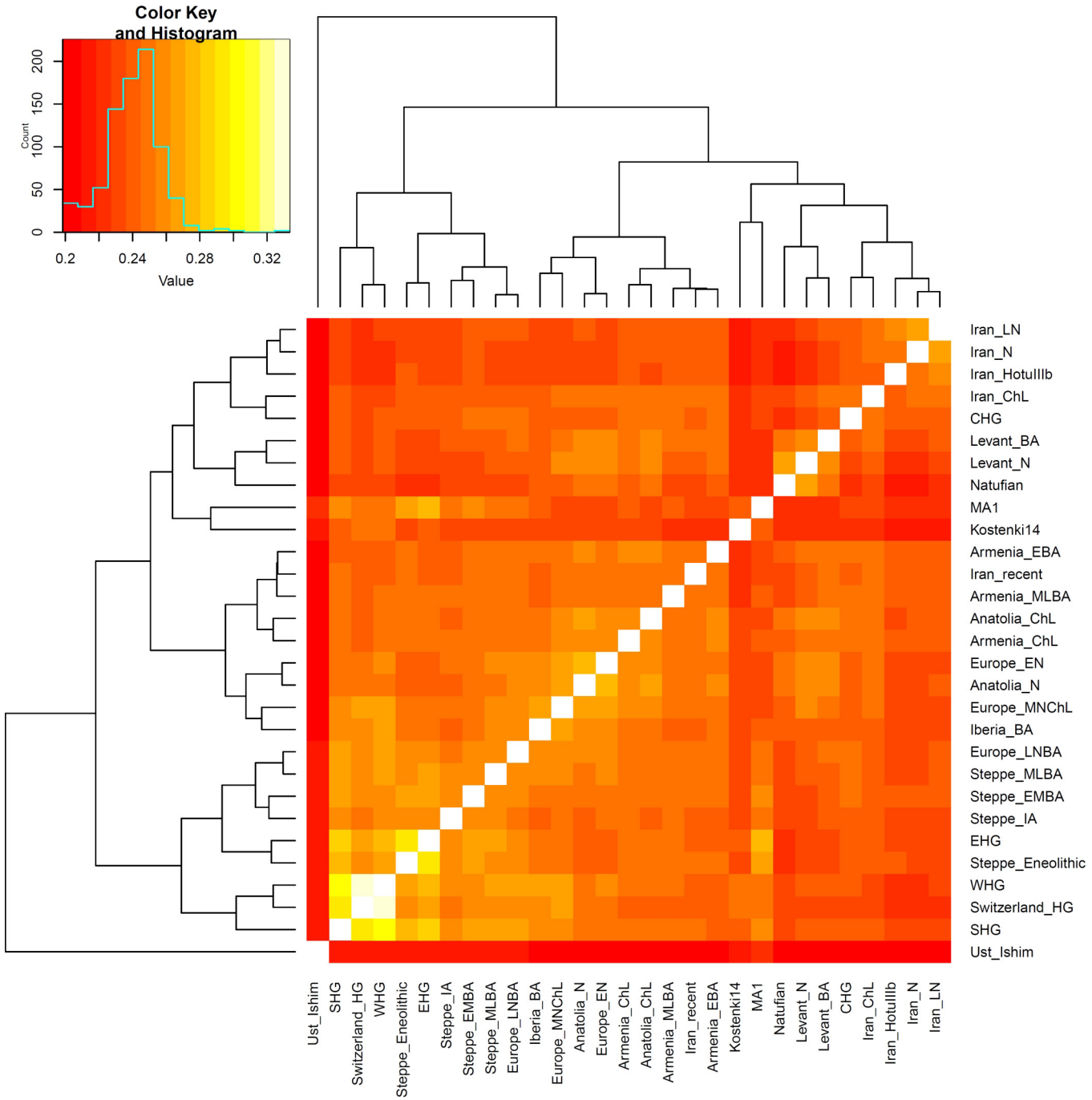
Extended Data Figure 1 | Principal components analysis of 991 present-day West Eurasians. The PCA analysis is performed on the same set of individuals as are reported in Fig. 1b, using EIGENSOFT. Here, we colour the samples by population (to highlight the present-day populations) instead of using grey points as in Fig. 1b (where the goal is to highlight ancient samples).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



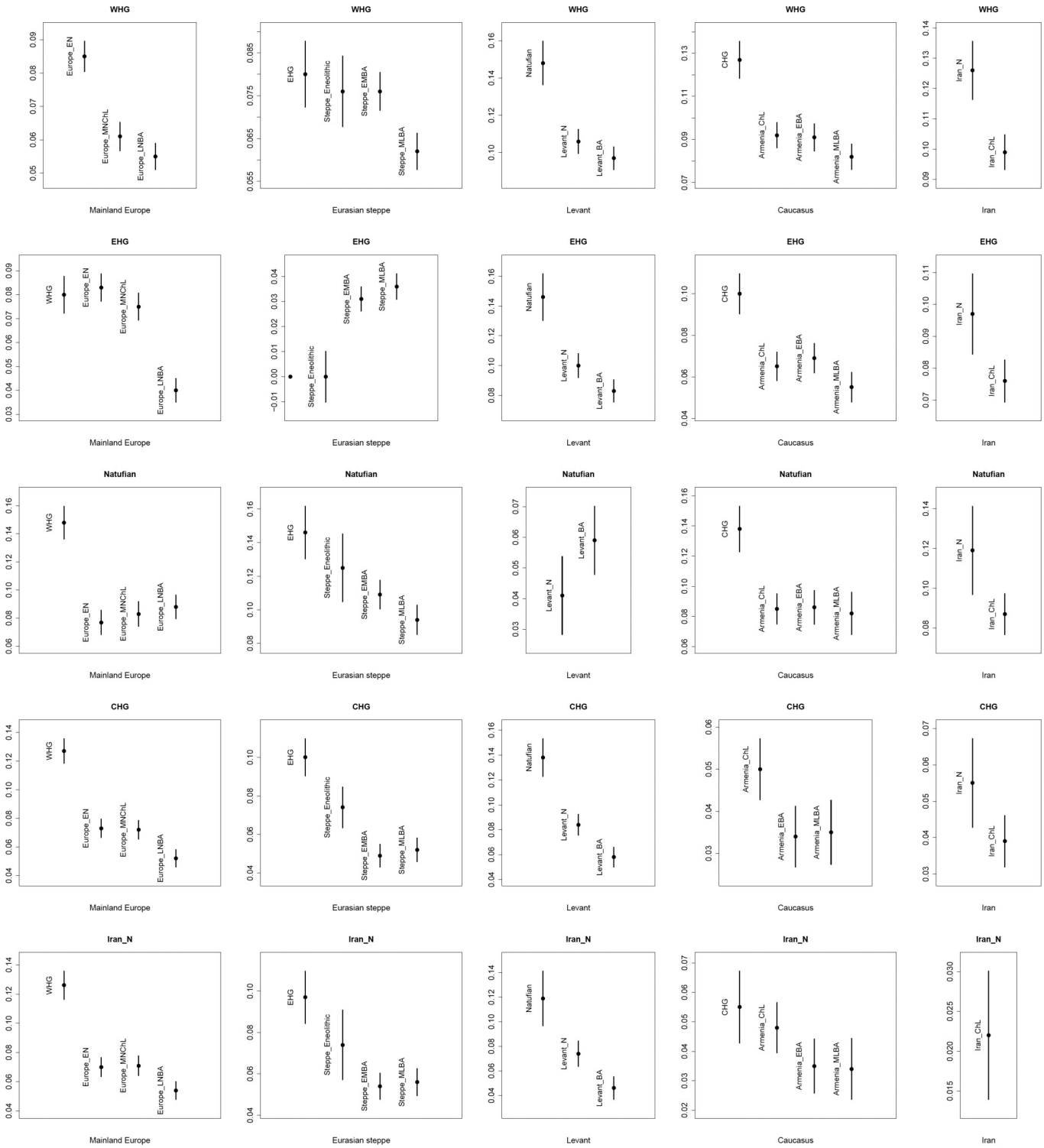
Extended Data Figure 2 | Genetic structure in ancient West Eurasian populations across time and decline of genetic differentiation over time. a, ADMIXTURE model-based clustering analysis of 2,583 present-day humans and 281 ancient samples; we show the results only for ancient samples for $K = 11$ clusters. **b**, Pairwise F_{ST} between 19 Ancient West Eurasian populations (arranged in approximate chronological order), and select present-day populations.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



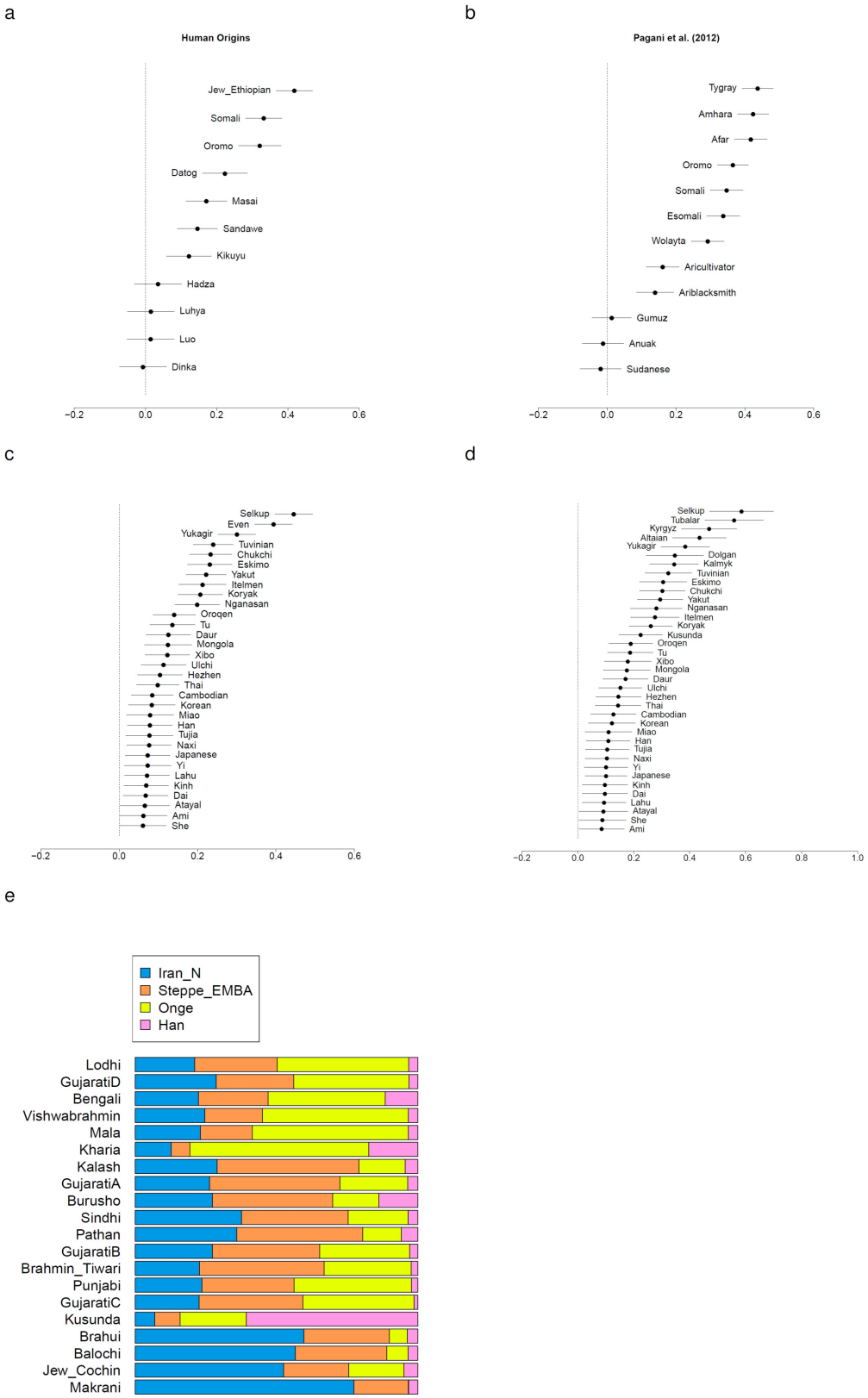
Extended Data Figure 3 | Outgroup f_3(Mbuti; X, Y) for pairs of ancient populations. The dendrogram is plotted for convenience and should not be interpreted as a phylogenetic tree. Areas of high shared genetic drift are 'yellow' and include from top-right to bottom-left along the diagonal: early Anatolian and European farmers; European hunter–gatherers, Steppe populations and populations admixed with steppe ancestry; populations from the Levant from the Epipalaeolithic (Natufians) to the Bronze Age; populations from Iran from the Mesolithic to the Late Neolithic.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 4 | Reduction of genetic differentiation in West Eurasia over time. We measure differentiation by F_{ST}. Each column of the 5 × 5 matrix of plots represents a major region and each row the earliest population with at least two individuals from each major region.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



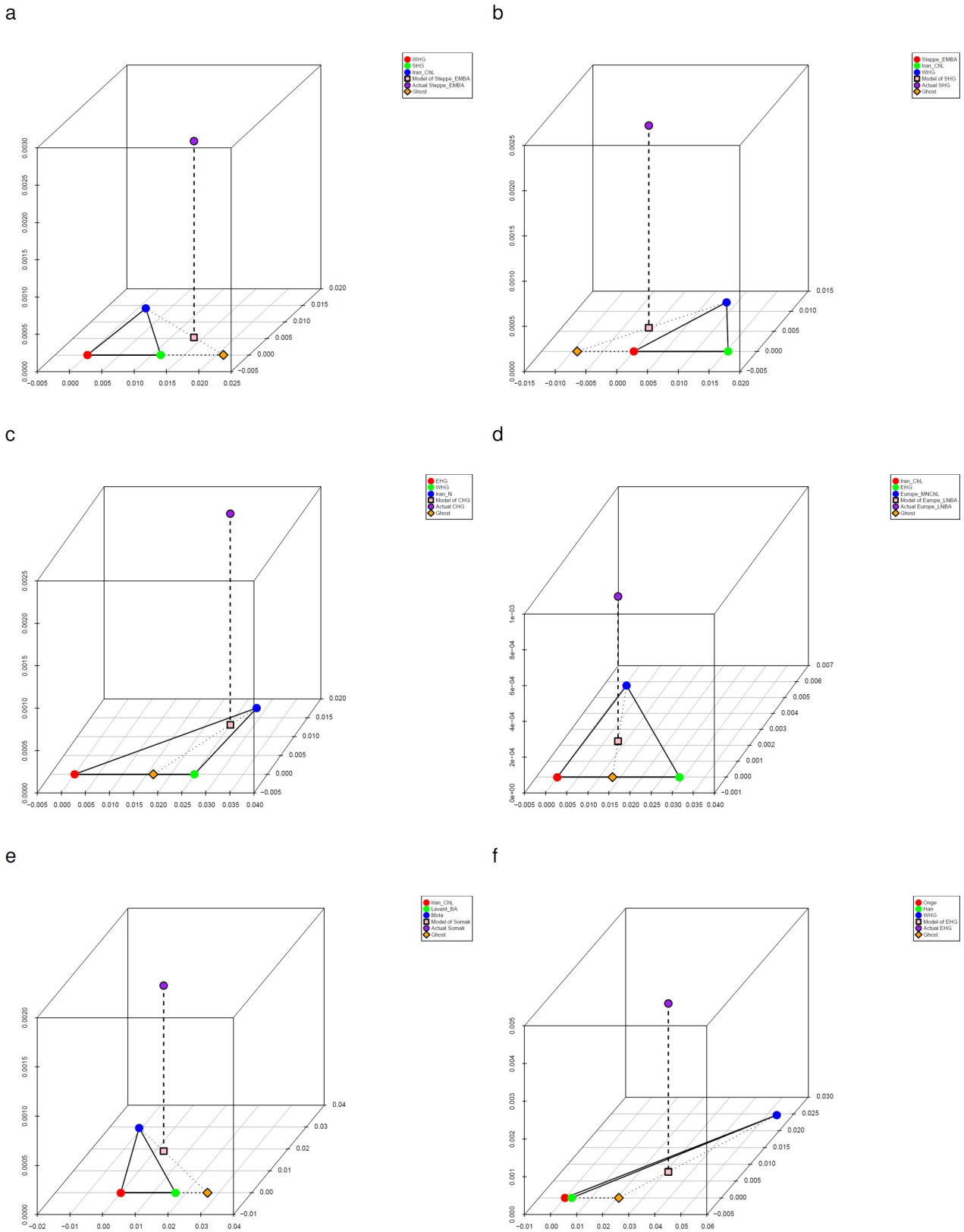
Extended Data Figure 5 | West Eurasian related admixture in East Africa, Eastern Eurasia and South Asia. a, Levantine ancestry in Eastern Africa in the Human Origins dataset. **b**, Levantine ancestry in different Eastern African population in the dataset from Pagani *et al.* (2012); the remainder of the ancestry is a clade with Mota, a ~4,500 year old sample from Ethiopia[49]. **c**, EHG ancestry in Eastern Eurasians. **d**, Afontova Gora (AG2)-related ancestry in Eastern Eurasians; the remainder of their ancestry is a clade with Onge. **e**, Mixture proportions for South Asian populations showing that they can be modelled as having West Eurasian-related ancestry similar to that in populations from both the Eurasian steppe and Iran.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



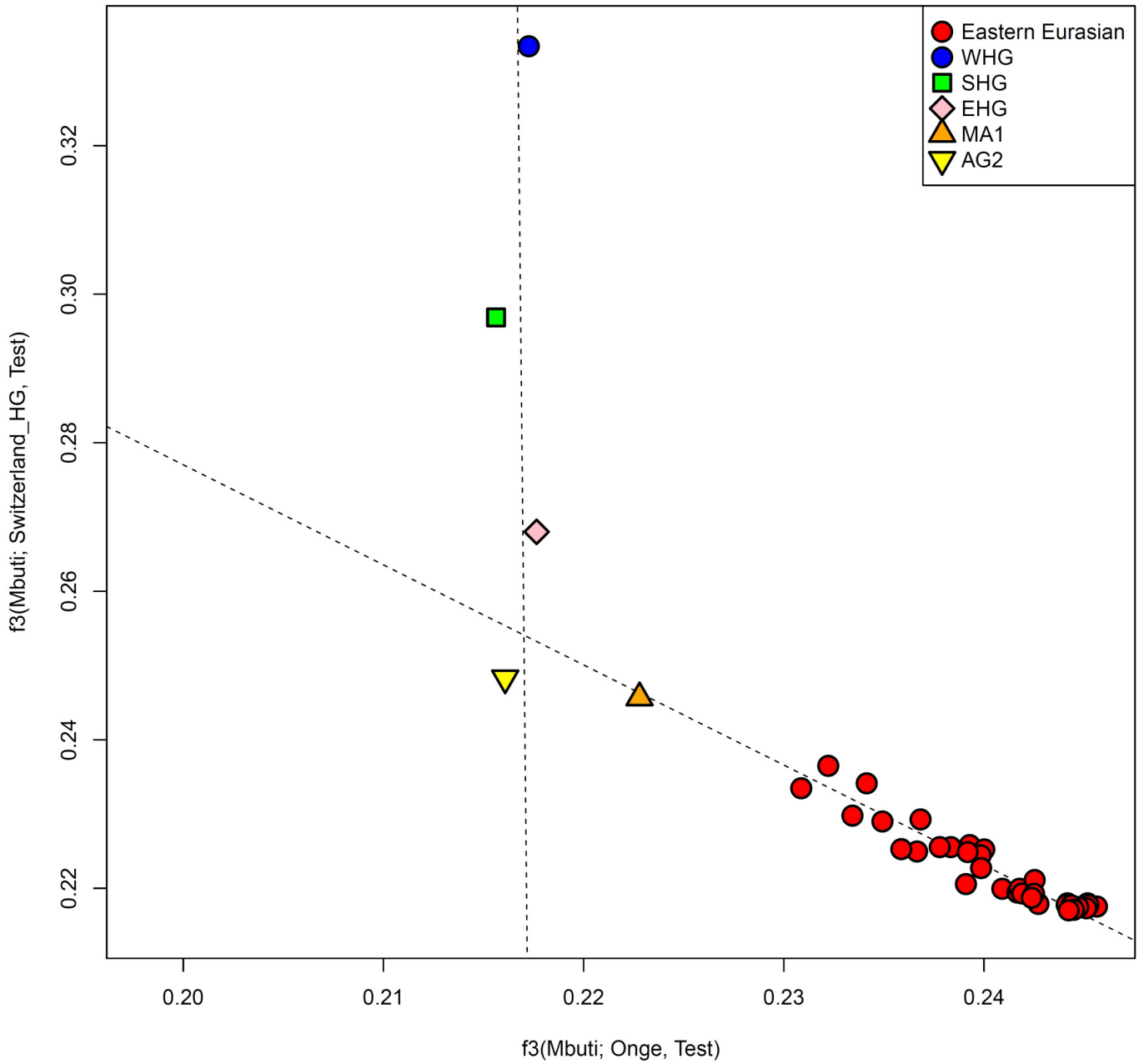
Extended Data Figure 6 | Inferred position of ancient populations in West Eurasian PCA according to the model of Fig. 4.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 7 | Admixture from ghost populations using 'cline intersection'. a–**f**, We model each *Test* population (purple) as a mixture (pink) of a fixed reference population (blue) and a ghost population (orange) residing on the cline defined by two other populations (red and green) according to the visualization method of Supplementary Information, section 10. **a**, Early/Middle Bronze Age steppe populations are a mixture of Iran_ChL and a population on the WHG→SHG cline. **b**, Scandinavian hunter–gatherers (SHG) are a mixture of WHG and a population on the Iran_ChL→Steppe_EMBA cline. **c**, Caucasus hunter–gatherers (CHG) are a mixture of Iran_N and both WHG and EHG. **d**, Late Neolithic/Bronze Age Europeans are a mixture of the preceding Europe_MNChL population and a population with both EHG and Iran_ChL ancestry. **e**, Somali are a mixture of Mota[49] and a population on the Iran_ChL→Levant_BA cline. **f**, Eastern European hunter–gatherers (EHG) are a mixture of WHG and a population on the Onge→Han cline.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 8 | Admixture from a 'ghost' ANE population into both European and Eastern Eurasian ancestry. EHG, and Upper Palaeolithic Siberians Mal'ta 1 (MA1) and Afontova Gora 2 (AG2) are positioned near the intersection of clines formed by European hunter–gatherers (WHG, SHG, EHG) and Eastern non-Africans in the space of outgroup f_3-statistics of the form f_3(Mbuti; Papuan, *Test*) and f_3(Mbuti; Switzerland_HG, *Test*).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 1 | No evidence for admixture related to sub-Saharan Africans in Natufians

Other Ancient	*African*	f_4(Natufian, *Other Ancient*; *African*, Chimp)	Z	Number of SNPs
EHG	Mbuti	-0.00044	-1.0	254033
EHG	Yoruba	0.00029	0.7	254033
EHG	Ju_hoan_North	-0.00015	-0.4	254033
EHG	Mota	-0.00022	-0.4	253986
WHG	Mbuti	-0.00067	-1.7	261514
WHG	Yoruba	-0.00045	-1.1	261514
WHG	Ju_hoan_North	-0.00046	-1.2	261514
WHG	Mota	-0.00129	-2.3	261461
SHG	Mbuti	-0.00076	-2.0	255686
SHG	Yoruba	-0.00039	-1.0	255686
SHG	Ju_hoan_North	-0.00052	-1.4	255686
SHG	Mota	-0.00091	-1.7	255641
Switzerland_HG	Mbuti	-0.00018	-0.4	261322
Switzerland_HG	Yoruba	0.00019	0.4	261322
Switzerland_HG	Ju_hoan_North	0.00009	0.2	261322
Switzerland_HG	Mota	-0.00062	-0.9	261276
Kostenki14	Mbuti	0.00034	0.7	246765
Kostenki14	Yoruba	0.00120	2.3	246765
Kostenki14	Ju_hoan_North	0.00069	1.4	246765
Kostenki14	Mota	0.00036	0.5	246719
MA1	Mbuti	-0.00038	-0.7	191819
MA1	Yoruba	0.00009	0.2	191819
MA1	Ju_hoan_North	-0.00010	-0.2	191819
MA1	Mota	-0.00038	-0.5	191782
CHG	Mbuti	-0.00051	-1.2	261505
CHG	Yoruba	-0.00012	-0.3	261505
CHG	Ju_hoan_North	-0.00013	-0.3	261505
CHG	Mota	-0.00042	-0.7	261456
Iran_N	Mbuti	-0.00018	-0.4	232927
Iran_N	Yoruba	0.00036	0.8	232927
Iran_N	Ju_hoan_North	0.00041	0.9	232927
Iran_N	Mota	0.00006	0.1	232880

We computed the statistic f_4(Natufian, *Other Ancient*; *African*, Chimp) varying *African* to be Mbuti, Yoruba, Ju_hoan_North, or the ancient Mota individual. Gene flow between Natufians and African populations would be expected to bias these statistics positive. However, we find most of them to be negative in sign and all of them to be non-significant ($|Z| < 3$), providing no evidence that Natufians differ from other ancient samples with respect to African populations.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 2 | Admixture f_3-statistics

Test	*Reference*$_1$	*Reference*$_2$	f_3(*Test*; *Reference*$_1$, *Refrence*$_2$)	Z-score	Number of SNPs
Anatolia_N	Iberia_BA	Levant_N	-0.00034	-0.2	111632
Armenia_ChL	EHG	Levant_N	-0.00249	-1.5	167020
Armenia_EBA	Anatolia_N	CHG	-0.01017	-7.9	195596
Armenia_MLBA	Anatolia_N	Steppe_EMBA	-0.00809	-7.3	203796
CHG	Anatolia_ChL	Iran_HotuIIIb	0.02612	3.6	9884
EHG	Steppe_Eneolithic	Switzerland_HG	-0.00282	-0.9	67938
Europe_EN	Anatolia_N	WHG	-0.00494	-11.2	380684
Europe_LNBA	Europe_MNChL	Steppe_EMBA	-0.00920	-41.8	414782
Europe_MNChL	Anatolia_N	WHG	-0.01351	-26.8	363672
Iran_ChL	Anatolia_N	Iran_N	-0.01285	-10.6	167941
Iran_N	Iran_LN	Gana	-0.00462	-1.1	17804
Levant_BA	Iran_N	Levant_N	-0.00853	-4.7	118269
Levant_N	Europe_MNChL	Natufian	-0.00671	-3.6	61845
Natufian	Iberia_BA	Iran_HotuIIIb	0.07613	3.4	1054
SHG	Steppe_Eneolithic	Switzerland_HG	0.00728	3.2	154825
Steppe_EMBA	EHG	Abkhasian	-0.00756	-11.2	349359
Steppe_Eneolithic	EHG	Iran_LN	-0.01637	-4.2	25100
Steppe_MLBA	Europe_MNChL	Steppe_EMBA	-0.00573	-18.0	378298
WHG	Switzerland_HG	Saudi	-0.01562	-7.7	218758
Abkhasian	CHG	Sardinian	-0.00754	-13.1	387956
Adygei	Anatolia_N	Eskimo	-0.00699	-14.4	413128
Albanian	Europe_EN	Burusho	-0.00650	-16.8	395851
Armenian	Anatolia_N	Sindhi	-0.00603	-19.5	406021
Assyrian	Iran_N	Sardinian	-0.00672	-11.8	309055
Balkar	Anatolia_N	Chukchi	-0.00975	-18.8	401928
Basque	Switzerland_HG	Druze	-0.00726	-12.6	416070
BedouinA	Europe_EN	Yoruba	-0.01584	-42.8	460762
BedouinB	Iran_HotuIIIb	Natufian	0.01384	4.1	32266
Belarusian	WHG	Iranian	-0.00974	-19.8	392363
Bulgarian	Anatolia_N	Steppe_EMBA	-0.00807	-26.7	400263
Canary_Islander	Europe_MNChL	Mende	-0.00829	-5.9	353172
Chechen	Anatolia_N	Eskimo	-0.00440	-7.9	396678
Croatian	WHG	Druze	-0.00871	-18.6	394032
Cypriot	Anatolia_N	Sindhi	-0.00562	-16.1	401141
Czech	SHG	Druze	-0.00919	-21.7	374705
Druze	Iran_N	Sardinian	-0.00269	-5.8	343813
English	Steppe_EMBA	Sardinian	-0.00628	-20.6	402502
Estonian	SHG	Druze	-0.00789	-17.6	371575
Finnish	SHG	Assyrian	-0.00716	-12.6	355744
French	Steppe_EMBA	Sardinian	-0.00669	-37.9	441807
Georgian	CHG	Sardinian	-0.00782	-13.7	390744
German	WHG	Druze	-0.01103	-22.9	391302
Greek	Europe_EN	Pathan	-0.00600	-30.0	421984
Hungarian	Steppe_EMBA	Sardinian	-0.00644	-31.2	420017
Icelandic	WHG	Abkhasian	-0.00974	-17.0	394625
Iranian	Anatolia_N	Sindhi	-0.00594	-30.9	443011
Irish	Steppe_EMBA	Sardinian	-0.00590	-22.8	416663
Irish_Ulster	SHG	Assyrian	-0.00909	-15.6	350547
Italian_North	Europe_EN	Steppe_EMBA	-0.00627	-26.4	419169
Italian_South	Iberia_BA	Iran_HotuIIIb	0.01224	2.6	17678
Jew_Ashkenazi	Anatolia_N	Koryak	-0.00532	-9.4	389012
Jew_Georgian	Iran_N	Sardinian	-0.00306	-4.2	292410
Jew_Iranian	Iran_N	Sardinian	-0.00385	-5.8	302446
Jew_Iraqi	Iran_N	Sardinian	-0.00486	-6.5	287673
Jew_Libyan	Europe_EN	Yoruba	-0.00397	-7.2	415797
Jew_Moroccan	Europe_EN	Yoruba	-0.00649	-10.9	405193
Jew_Tunisian	Anatolia_N	Mende	-0.00276	-4.1	399354
Jew_Turkish	Anatolia_N	Burusho	-0.00571	-16.4	405254
Jew_Yemenite	Natufian	Kalash	-0.00341	-3.8	174052
Jordanian	Europe_EN	Yoruba	-0.01283	-26.7	423649
Kumyk	Anatolia_N	Chukchi	-0.01025	-19.6	396439
Lebanese	Anatolia_N	Yoruba	-0.01022	-19.5	414854
Lebanese_Christian	Anatolia_N	Sindhi	-0.00504	-15.7	404858
Lebanese_Muslim	Anatolia_N	Brahmin_Tiwari	-0.00616	-20.4	415129
Lezgin	Steppe_EMBA	Jew_Yemenite	-0.00481	-13.1	398974
Lithuanian	WHG	Abkhasian	-0.00999	-17.7	386718
Maltese	Anatolia_N	Brahmin_Tiwari	-0.00518	-14.5	404438
Mordovian	WHG	Iranian	-0.00912	-18.4	395230
North_Ossetian	Anatolia_N	Chukchi	-0.00894	-17.2	401729
Norwegian	WHG	Abkhasian	-0.00957	-16.5	393546
Orcadian	SHG	Druze	-0.00662	-15.8	379656
Palestinian	Europe_EN	Yoruba	-0.01129	-31.3	464066
Polish	SHG	Druze	-0.00924	-27.8	394654
Romanian	Europe_EN	Steppe_EMBA	-0.00549	-16.9	397119
Russian	SHG	Turkish	-0.00731	-25.0	398393
Sardinian	Anatolia_N	Switzerland_HG	-0.00587	-9.6	417931
Saudi	Anatolia_N	Dinka	-0.00326	-5.1	404923
Scottish	Steppe_EMBA	Sardinian	-0.00622	-26.6	426660
Shetlandic	WHG	Abkhasian	-0.00868	-14.6	386562
Sicilian	Anatolia_N	Brahmin_Tiwari	-0.00646	-22.2	411481
Sorb	SHG	Palestinian	-0.00787	-16.8	366924
Spanish	Steppe_EMBA	Sardinian	-0.00557	-32.2	447735
Spanish_North	WHG	Armenian	-0.00825	-10.9	356832
Syrian	Europe_EN	Dinka	-0.01002	-17.3	410920
Turkish	Europe_EN	Sindhi	-0.00709	-41.1	448975
Ukrainian	WHG	Abkhasian	-0.01183	-21.4	388282

We show the lowest *Z*-score of the statistic f_3(*Test*; *Reference*$_1$, *Refrence*$_2$) for *Test* populations with at least 2 individuals and every pair (*Reference*$_1$, *Refrence*$_2$) of ancient or present-day source populations. *Z*-scores lower than −3 are highlighted and indicate that the *Test* population is admixed from sources related to (but not identical to) the reference populations. *Z*-scores greater than −3 are consistent with the population either being admixed or not.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

ARTICLE

doi:10.1038/nature19094

Uncovering Earth's virome

David Paez-Espino[1], Emiley A. Eloe-Fadrosh[1], Georgios A. Pavlopoulos[1], Alex D. Thomas[1], Marcel Huntemann[1], Natalia Mikhailova[1], Edward Rubin[1,2,3], Natalia N. Ivanova[1] & Nikos C. Kyrpides[1]

Viruses are the most abundant biological entities on Earth, but challenges in detecting, isolating, and classifying unknown viruses have prevented exhaustive surveys of the global virome. Here we analysed over 5 Tb of metagenomic sequence data from 3,042 geographically diverse samples to assess the global distribution, phylogenetic diversity, and host specificity of viruses. We discovered over 125,000 partial DNA viral genomes, including the largest phage yet identified, and increased the number of known viral genes by 16-fold. Half of the predicted partial viral genomes were clustered into genetically distinct groups, most of which included genes unrelated to those in known viruses. Using CRISPR spacers and transfer RNA matches to link viral groups to microbial host(s), we doubled the number of microbial phyla known to be infected by viruses, and identified viruses that can infect organisms from different phyla. Analysis of viral distribution across diverse ecosystems revealed strong habitat-type specificity for the vast majority of viruses, but also identified some cosmopolitan groups. Our results highlight an extensive global viral diversity and provide detailed insight into viral habitat distribution and host–virus interactions.

Viruses are the most abundant entities across all habitats, and a major reservoir of genetic diversity[1] affecting biogeochemical cycles and ecosystem dynamics[1]. Exploration of viral populations in oceans of the world and within the human microbiome has illuminated considerable genetic complexity[2,3]; however, there are significant gaps in the global virome catalogue. There are an estimated 10^{31} viral particles infecting microbial populations[4]; yet fewer than 2,200 genomes from double-stranded DNA (dsDNA) viruses and retroviruses are deposited in NCBI, compared to over 45,000 bacterial genomes[5]. Culture-independent approaches have provided a broader view of the diversity and distribution of dsDNA viruses[6]. However, their accurate detection and quantification using targeted sequencing remains challenging owing to the lack of universally conserved genomic signatures and complex experimental protocols[7].

Beyond gaps in characterized diversity, the scope of host–viral interactions is poorly understood, although it has been hypothesized that all cellular organisms are prey to viral attack[8]. Methods for studying host–viral interactions rely almost exclusively on cultured virus–host systems; however, recent *in silico* approaches have revealed that a much broader range of hosts is susceptible to viral infections[9,10]. Given the role that viruses play in host metabolism reprogramming, gene flow, and structuring of microbial communities, it is critical to capture viral linkages with their hosts.

Currently, a plethora of metagenomic data exists that present a unique opportunity for viral sequence discovery[11]. Although most of these data sets were generated by untargeted approaches without viral particle enrichment, they contain a wealth of viral sequences. Here, we developed a computational approach to explore the viral content of more than 3,000 metagenomic samples. We uncovered 2.1 Gb of viral sequence data, which increases the known viral sequence space by an order of magnitude, enables the prediction of previously unknown host–viral interactions and provides a global view of viral biogeography.

Global expansion of viral sequence space

In the absence of universally conserved markers, previous studies attempted to identify viruses using proteins present exclusively in viruses[12]. To overcome the limitations of a biased collection of isolate viruses (iVGs), we complemented the viral protein families of the iVGs (derived from dsDNA viruses and retroviruses in the NCBI database) with a set of viral protein families from 1,800 manually identified metagenomic viral contigs (mVCs). This set was used as a bait to identify putative viral sequences in a large collection of assembled metagenomic contigs longer than 5 kb (Methods; Extended Data Figs 1–3; Supplementary Tables 1–7). These contigs were obtained from 3,042 metagenomes in the Integrated Microbial Genomes with Microbiome Samples (IMG/M) system[11] (Supplementary Table 8), representing a collection of geographically and ecologically diverse samples according to metadata from the Genomes OnLine Database[5,13]. This led to the identification of 125,842 putative DNA metagenomic viral contigs, increasing the viral sequence size in base pairs by 17.3-fold and the number of viral genes by 16.6-fold (Fig. 1a; Methods). These encode more than 2.79 million proteins, 75% of which have no sequence similarity to proteins from known isolate viruses, consistent with previous studies[12,14]. Sequence similarity clustering of proteins encoded by the mVCs resulted in a total of 418,541 clusters with 2 or more members and 765,991 singletons (Methods; Supplementary Table 9). Benchmarking was performed to validate our computational pipeline, and indicated that 70% of the sequences identified in this study would have been missed by other methods (Methods; Extended Data Fig. 3; Supplementary Tables 2–6).

To evaluate the coverage of the viral protein space by the newly identified sequences, we estimated the rate of accumulation of protein clusters as a function of the number of samples (Fig. 1b). In agreement with recent reports[9,15], the curves of cluster accumulation in the two most heavily sampled habitats, human-associated and marine, appear to reach saturation. However, the rate of cluster discovery does not plateau when all samples are considered, suggesting that the global viral sequence space is largely uncharacterized.

To compare the coverage of mVCs and iVGs by viral protein families, we calculated the percentage of genes with hits to viral protein families relative to the total number of genes on each contig (Fig. 1c). On the basis of this percentage, viral contigs were classified into three categories: those with at least 70% of genes in viral families (highly covered with strong similarity to viruses in the training set); those with 35–70% of genes in viral families; and those with less than 35% of genes in viral families (low covered with low similarity to viruses

[1]Department of Energy, Joint Genome Institute, Walnut Creek, California 94598, USA. [2]Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. [3]Metabiota Inc., San Francisco, California 94104, USA.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

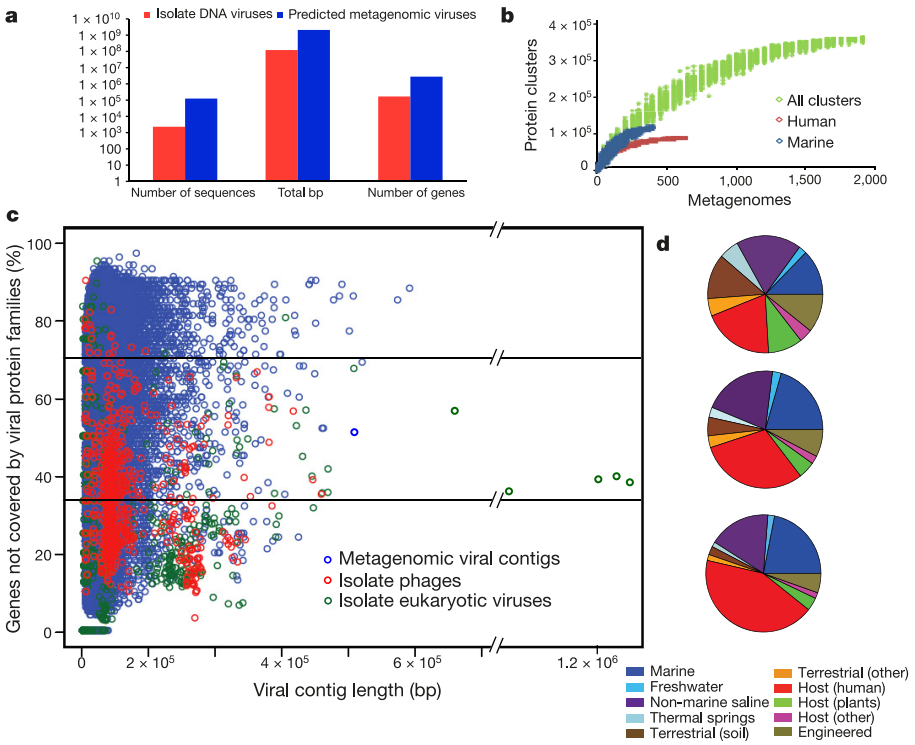


Figure 1 | Identification of metagenomic viral sequences and habitat distribution. a, Number of metagenomic viral contigs compared to isolate viral genomes. **b**, Accumulation curves showing the protein cluster growth with increased sampling. Green, blue, and red represent clusters from all, aquatic, or human metagenomes, respectively. The ranges represent 100 random replicates. **c**, Distribution of metagenomic viral contigs (blue), isolate phages (red) and isolate eukaryotic viruses (green) by length and coverage with viral protein families. Viral sequences are divided into three categories: high, medium and low coverage with protein families. **d**, Distribution of metagenome viral contigs by habitat type is shown.

in the training set) (Extended Data Fig. 4a). The highly covered category included 67% of isolate viruses, but only 24.5% of mVCs (Fig. 1c), the majority of them from marine and human-associated habitats, where more reference viruses were available (Fig. 1d). Another 24.2% of mVCs placed in the low-covered category were typically found in soil, plant-associated, and engineered samples (Fig. 1d). The differences were even more pronounced when the data were normalized by total sequence length per habitat (Methods), suggesting the need for more extensive sampling of these environments.

The length of mVCs ranges from 5 kb to nearly 600 kb (average 16,625 ± 18,057 bp) (Fig. 1c). On the basis of the end overlaps, 999 of mVCs were probably circular, representing complete viral genomes (Supplementary Table 10). The average size of the circular mVCs (53,644 ± 45,677 bp) is consistent with the calculated average length of isolate dsDNA viruses (44,296 ± 83,777 bp) (Supplementary Information). Among circular contigs, we identified the largest phage recovered to date, a 596 kb contig from a bioreactor sample[16], with many signature genes of tailed viruses, but no recognizable housekeeping genes of bacteria or plasmids (Methods; Extended Data Fig. 4b; Supplementary Table 11; Supplementary Information). We identified six more mVCs ranging from 350 to 470 kb, probably representing fragments of other large phage genomes (Supplementary Table 12). As the sizes of viral particles and viral genomes are correlated[17], these mVCs found in many ecological niches point to the hidden diversity and abundance of very large phages, probably avoiding detection by the conventional enrichment methods.

Sequence grouping to gauge viral diversity

To quantify the amount of taxonomic diversity, mVCs and isolate viral genomes were clustered into quasi-species groups on the basis of the average amino acid identity[18] (AAI) of all proteins and single-linkage clustering, using an approach analogous to the whole-genome-based classification scheme developed for prokaryotes[19] (Methods; Extended Data Fig. 5a; Supplementary Table 13). 64,160 mVCs and 2,536 isolate contigs were clustered into 18,470 viral groups, ranging from 2 to 365 members per group. Most groups (57%) had only 2 members and only 3.7% had more than 10 members (Extended Data Fig. 5b). Similar to previous studies[12,14,20], the vast majority of viral groups (95.9%) did not contain isolate viruses.

218 viral groups and 842 singletons contained at least one iVG with genus- and species-level taxonomic assignment according to the International Committee on Taxonomy of Viruses (Methods; Supplementary Tables 14, 15). Our method recapitulates current species-level groupings in 87% of the cases with the remainder grouping at genus level (Supplementary Table 14; Extended Data Fig. 5c–e). We compared our method with sequence-based classification used in previous studies, which applied protein cluster occurrence to generate mostly genus-level groups[21]. In agreement with an assessment that our groups represent quasi-species, our approach resulted in smaller clusters and more singletons (Supplementary Information). Next we proceeded to predict host specificity and determine environmental distribution of these species-level viral groups and singletons.

Host–virus connectivity revealed

We used a suite of computational methods to identify putative host–virus connections. First, we projected the isolate viral–host information onto a group, resulting in host assignments for 2.4% of viral groups (Fig. 2a). Then we used the CRISPR–Cas prokaryotic immune system, which holds a 'library' of genome fragments from phages (proto-spacers) that have previously infected the host[22]. These fragments retained by the host in the form of spacers can be matched to phage genomes linking phages with their hosts[23–25]. We amassed a database of 3.5 million spacers from prokaryotic isolate genomes and metagenomes in IMG (Supplementary Tables 16, 17). As a control, 98.5% of spacer matches against isolate viral genomes agreed with their known host specificity at genus or species level (Methods; Supplementary Information; Supplementary Table 18). Spacers from isolate microbial genomes with matches to mVCs were identified for 4.4% of the viral

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

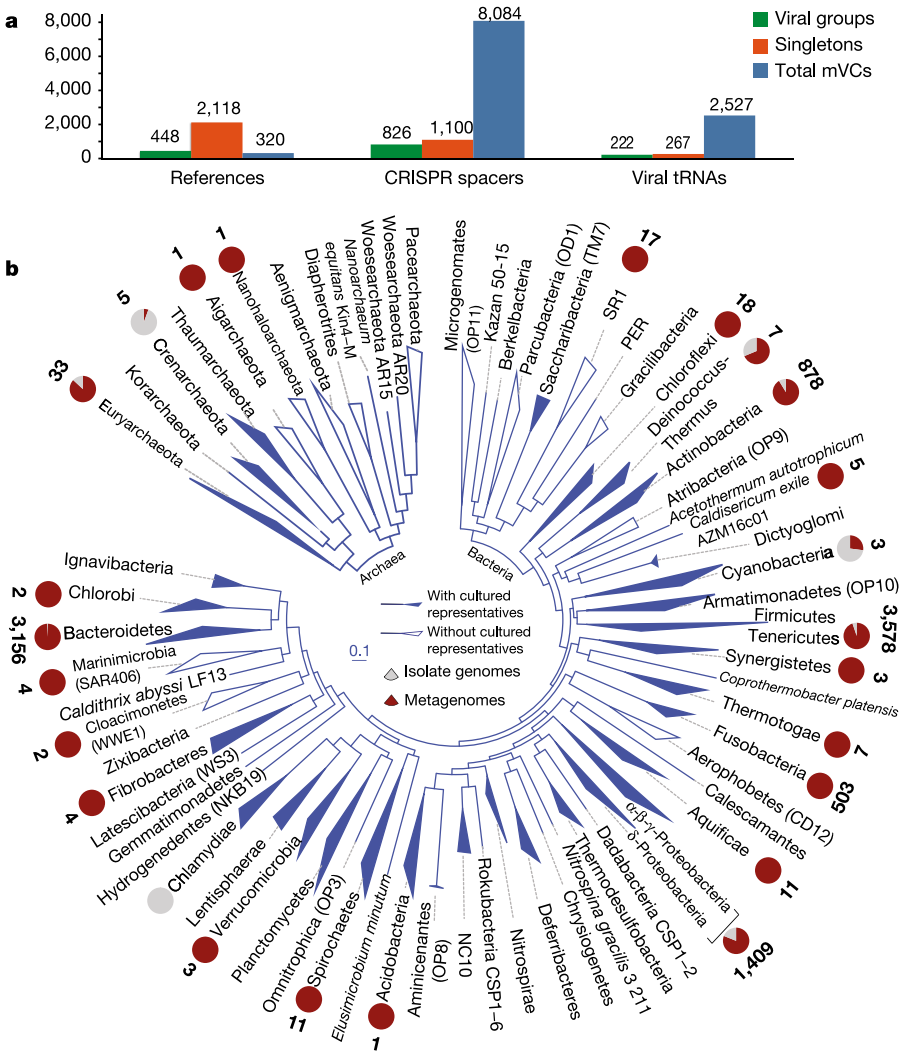


Figure 2 | Host–virus connectivity. a, Total number of host assignments to metagenomic viruses with three approaches. Total assignments for viral groups, singletons, and metagenomic viral contigs are shown above each bar. **b**, Phylogenetic distribution of bacterial and archaeal hosts. For each phylum a pie chart indicates the fraction of sequences assigned to this phylum from metagenomic viral contigs (red), and isolate viruses (grey). The number of metagenomic viral contigs assigned to each phylum is indicated by the numbers next to pie charts. Clades in blue represent phyla with cultivated representatives. Clades in white represent candidate phyla without cultured representatives.

groups and 1.7% of singletons (Supplementary Table 19). Finally, we explored the hypothesis that viral transfer RNA (tRNA) genes originate from their host[26]. Using stringent sequence identity cutoffs, viral tRNAs identified in 7.6% of the mVCs were matched to isolate genomes from a single species or genus (Methods; Supplementary Information; Supplementary Tables 20–22). The specificity of tRNA-based host-viral assignment was confirmed by CRISPR–Cas spacer matches showing a 94% agreement at the genus level (Supplementary Table 19).

Overall, these approaches identified 9,992 putative host–virus associations enabling host assignment to 7.7% of mVCs. The majority of these connections were previously unknown, and include hosts from 16 prokaryotic phyla for which no viruses have previously been identified (Supplementary Table 23), such as the first instance of viruses infecting the candidate phylum SR1 (Fig. 2b). We also connected mVCs to pathogenic species for which no viral connections were known, including *Fusobacterium* and *Leptotrichia* that cause oral and skin infections in mammals (Supplementary Information; Supplementary Table 24). The discovery of phages infecting these and other pathogens could be exploited for phage therapy applications[27,28].

It is widely assumed that most viruses specialize in infecting related hosts, as broad host range is negatively correlated with infection success[29]. However, this may be an artefact[29], and viral generalists that infect hosts across taxonomic orders do exist[30]. Our data suggested a trend for narrow host range with some notable exceptions. Whereas most CRISPR spacer matches were from viral sequences to hosts within one species or genus (Fig. 3a), some mVCs were linked to multiple hosts from higher taxa, including different phyla. A viral group comprised of mVCs from human oral samples contained three distinct proto-spacers with nearly exact matches to spacers in Actinobacteria and Firmicutes (Fig. 3b). In another case (Fig. 3c), proto-spacers from two mVCs derived from faecal samples were linked to spacers in three distinct Clostridiales families (Extended Data Fig. 6; Supplementary Information). As viruses exploit the host transcription/translation machinery, the existence of viruses with a surprisingly broad range of hosts opens opportunities for identification of novel enzymes or regulatory sequences, with biotechnological applications.

Biogeographic patterns of viral diversity

Previous studies of viral biogeography mainly focused on single habitats[3,31,32], and only a handful of small-scale studies explored viromes across environments[33–35]. As our pipeline was designed to identify longer viral contigs that probably represent more abundant populations, we explored the dispersal of the predicted viruses by aligning the contigs against all assembled and unassembled

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

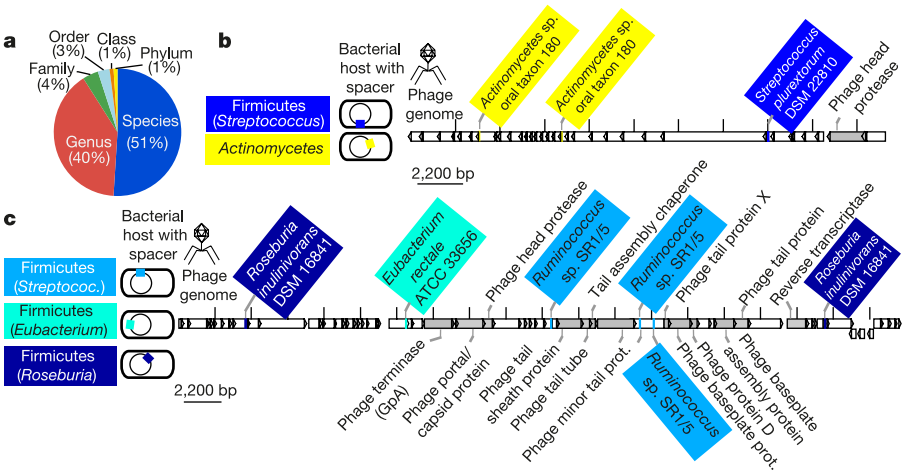


Figure 3 | Expanded host-range specificity identifies viral generalists. a, Proportion of viral groups connected with predicted hosts at various taxonomic levels. **b**, Three proto-spacers encoded on mVCs identified in human oral metagenomic samples that were linked to CRISPR spacers from hosts from distinct phyla, *Actinomycetes* sp. oral taxon 180 (Actinobacteria) and *Streptococcus plurextorum* DSM 22810 (Firmicutes). **c**, Seven proto-spacers encoded on two nearly identical mVCs deriving from independent faecal samples, and linked to CRISPR spacers from *Roseburia inulinivorans* DSM 16841 (Lachnospiraceae family), *Eubacterium rectale* ATCC 33656 (Eubacteriaceae family), and *Ruminococcus* sp. SR 1/5 (Ruminococcaceae family) (details in Supplementary Information).

metagenome sequences (Methods; Supplementary Information). This approach revealed that 86% of the viral sequences were found in more than one sample, whereas 73% were present in at least 5 samples (Extended Data Fig. 7a), mostly from relatively well-sampled marine and human-associated habitats. This enabled a detailed investigation of viral distribution patterns across these environments (Fig. 4).

The distribution of viral sequences in marine samples is characterized by distinct spatial patterns based on water column depth and distance from the shore (Fig. 4a). Although viral assemblages in coastal waters from distant biogeographical provinces are markedly different, bathypelagic samples from different oceanic basins display very similar viral profiles, in agreement with observations that deep ocean phylum-level composition of microbial communities is relatively uniform[36]. Although viral sequences are mostly partitioned into zone-specific groups, some are present in diverse samples across zones and oceanic provinces, including one viral group found in 95% of all twilight samples and in 44% of deep ocean samples (Extended Data Fig. 7b–c).

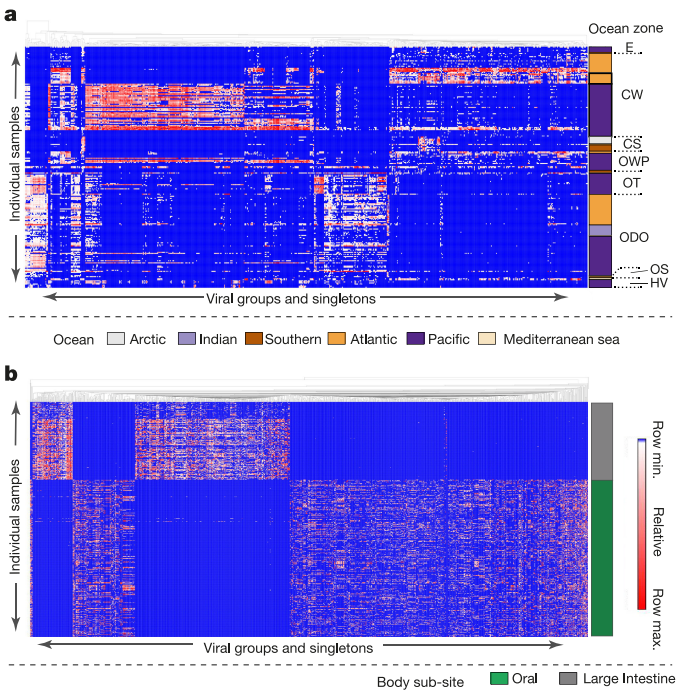


Figure 4 | Viral distribution patterns in marine and human samples. **a**, **b**, Hierarchical clustering of viral groups and singletons across marine (**a**) and human samples (**b**). Data sets were grouped according to environmental metadata or body sub-site, respectively. Oceanic zones (**a**) include Estuary (E), Coastal waters (CW), Coastal sediments (CS), Oceanic water photic (OWP; surface to 200m depth), Oceanic twilight (OT; 200m to 750m depth), Oceanic deep ocean (ODO; below 750m), Oceanic sediment (OS), and Hydrothermal vents (HV). Virus coverage is colour-coded from white (lowest coverage) to red (highest coverage) as shown in **b** inset. Blue represents absence of the corresponding viral sequence.

The distribution of viral sequences in human microbiome samples (Fig. 4b) also shows clear body-site specificity with only a few viral groups and singletons found in both faecal and oral samples (Supplementary Tables 25, 26). In contrast to previous studies[2,37–39], many viral sequences, mostly of phage origin, were shared between samples from the same body site of unrelated individuals. More than 30% of intestinal and 50% of oral viral sequences were shared by at least 10% of sampled subjects (Extended Data Fig. 7d, e). Approximately 0.5% of sequences in both body sites were shared by more than 80% of sampled individuals, whereas 17% and 9% of intestinal and oral viral sequences, respectively, were unique to each individual. We used raw sequencing reads to estimate the amount of viral sequences in 550 faecal and oral samples. Viral fraction varied from 0.2 to 54% of the total amount of high quality sequence in the sample, with the average of 3.4% in oral samples and 7.4% in stool samples, which is higher than previously reported 2.5 to 3.5% in stool[2] (Supplementary Information).

Although 84% of our quasi-species viral groups found in multiple samples resided within a single habitat type (Fig. 5a), 14% were found in two habitat types, typically, within the same broader environmental category (Fig. 5b, c), and a small number of groups were spanning two or more environmental categories (Fig. 5c, d; Supplementary Information; Supplementary Table 27). Most of these were due to uncertainty of habitat classification (for example, plant rhizosphere samples classified as host-associated) (Fig. 5c). A more detailed analysis of the most ubiquitous viral sequences revealed that they are probably human and laboratory contaminants, including ΦX and λ phages used as vectors, sequencing and molecular weight standards[40], and *Propionibacterium acnes* phages, common inhabitants of human skin. Several viruses recovered in a wide variety of environments were found to be prophages with broad host specificity (Supplementary Information; Supplementary Table 28) infecting hosts with different habitat preferences. Some of these prophages were found to carry a variety of cargo genes, presumably conferring competitive advantage to their hosts and explaining their broad distribution[41] (Extended Data Figs 8, 9). However, in a few cases, the presence of viral groups in diverse environments could not be attributed to metadata discrepancies, ambiguity of habitat classification, contamination or broad host specificity. A small number of viral groups was found in aquatic samples with large differences in salinity, such as freshwater and hypersaline lakes, whereas other groups were found in oil-contaminated wastewater, and in human oral and faecal samples (Fig. 5c, d; Supplementary Information). Our observations of a limited number of ubiquitous viruses expand on

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

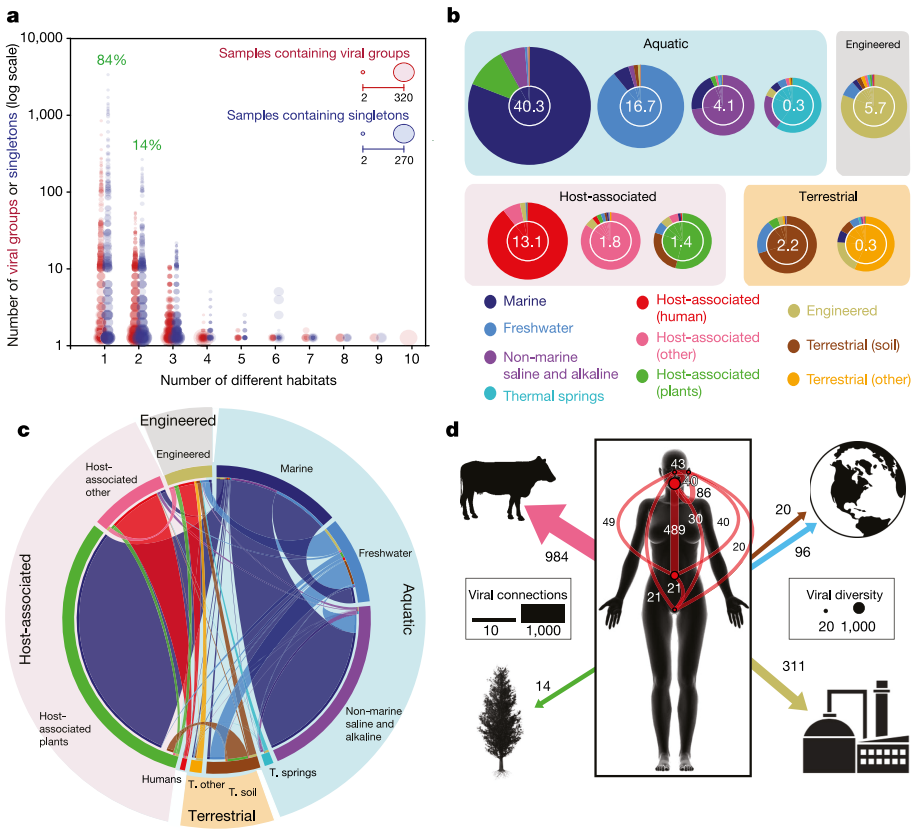


Figure 5 | Habitat distribution of metagenomic viruses. a, Distribution of viral groups and singletons (red and blue bubbles, respectively) against the number of habitats. Bubble size (inset) reflects the number of samples containing viral sequence. **b**, Distribution of viral sequences per habitat type and environmental category. Pie charts size and the number inside show viral groups and singletons per habitat type grouped into four environmental categories; values are given in units of 10^3. **c**, Pairwise connection between habitat types based on shared viral sequences; **d**, Distribution of viral sequences across human body sub-sites and across habitat types. Red dots indicate viral groups and singletons exclusive for each sub-site.

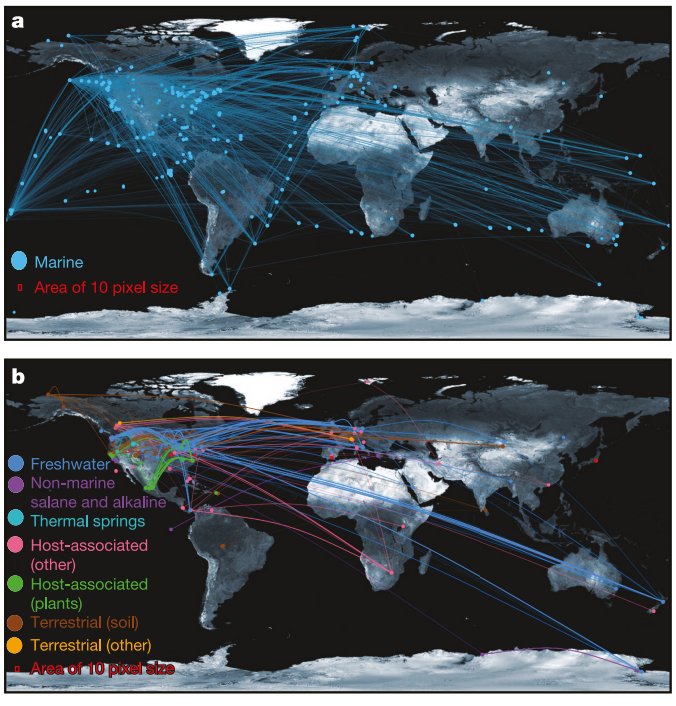


Figure 6 | Global distribution of viral diversity. The presence of the same viral groups or singletons across samples (circles) is represented by connecting lines. Only samples over 10 pixels apart and sharing 2 or more viral groups or singletons are shown. The colours of the circles indicate the habitat type. **a**, Connections between marine samples. Transparency of the lines reflects the number of shared viral groups. **b**, Connections between non-marine samples of the same habitat type. Connections between different habitat types are shown in Extended Data Fig. 10. Equirectangular projection of the world map is used as background image (http://visibleearth.nasa.gov/view.php?id=57752).

previous studies[35,42] that shed light on the mechanisms underlying their 'cosmopolitanism'.

To generate a global map of viral dispersal, we linked the viral sequences with geographic coordinates of the corresponding samples. Many viruses were found in similar ecological niches across large geographic distances, with the most prominent connectivity within extensively sampled marine biome (Fig. 6a; Supplementary Table 8), which is in agreement with previous studies, suggesting that viruses are passively transported along oceanic currents[3]. We also observed sparse but non-negligible connections between non-marine viral groups and singletons in samples of the same habitat type across biomes (Fig. 6b) and across different ecosystems (Extended Data Fig. 10).

Discussion

This study shows that in-depth exploration of ecosystems by untargeted metagenome sequencing is a powerful approach to fill knowledge gaps and address fundamental questions of viral ecology. Our analysis led to a notable increase in the number of viral sequences and putative virus–host connections, demonstrating that a much larger prokaryotic diversity than previously known is preyed upon by viruses, expanding on a recent report of microbial lineages containing prophages[21]. Consistent with previous observations, the environmental viral quasi-species were mostly found to have a narrow host range with a few notable exceptions of phages with broad taxonomic host specificity, including examples of hosts from different phyla.

The global maps of viral biogeography show that viruses are predominantly found in similar habitats, regardless their geographic proximity. This pattern was most prominent for marine viruses as previously observed[3], yet was also striking across seemingly isolated locales such as lakes, plant-associated habitats and soils, where the dispersal mode is not immediately obvious. More surprising was the significant human virome sharing between unrelated individuals and the identification of viral quasi-species distributed across markedly different habitats.

Overall, this study demonstrates the value of untargeted *de novo* metagenomic analysis as compared to reference-based and targeted

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

virome approaches, highlighting the importance of globally sampled metagenomic data sets[43] to vastly improve viral sequence discovery. Ultimately, large-scale computational exploration of uncharted viral sequence space will assist in addressing the remaining mysteries of viral ecology.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 November 2015; accepted 8 July 2016.
Published online 17 August 2016.

1. Suttle, C. A. Marine viruses—major players in the global ecosystem. *Nat. Rev. Microbiol.* **5,** 801–812 (2007).
2. Reyes, A. *et al.* Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* **466,** 334–338 (2010).
3. Brum, J. R. *et al.* Ocean plankton. Patterns and ecological drivers of ocean viral communities. *Science* **348,** 1261498 (2015).
4. Whitman, W. B., Coleman, D. C. & Wiebe, W. J. Prokaryotes: the unseen majority. *Proc. Natl Acad. Sci. USA* **95,** 6578–6583 (1998).
5. Reddy, T. B. *et al.* The Genomes OnLine Database (GOLD) v.5: a metadata management system based on a four level (meta)genome project classification. *Nucleic Acids Res.* **43,** D1099–D1106 (2015).
6. Chow, C. E. & Suttle, C. A. Biogeography of viruses in the sea. *Annu Rev Virol* **2,** 41–66 (2015).
7. Rohwer, F. & Edwards, R. The Phage Proteomic Tree: a genome-based taxonomy for phage. *J. Bacteriol.* **184,** 4529–4535 (2002).
8. Fuhrman, J. A. Marine viruses and their biogeochemical and ecological effects. *Nature* **399,** 541–548 (1999).
9. Brum, J. R. & Sullivan, M. B. Rising to the challenge: accelerated pace of discovery transforms marine virology. *Nat. Rev. Microbiol.* **13,** 147–159 (2015).
10. Edwards, R. A., McNair, K., Faust, K., Raes, J. & Dutilh, B. E. Computational approaches to predict bacteriophage-host relationships. *FEMS Microbiol. Rev.* **40,** 258–272 (2016).
11. Markowitz, V. M. *et al.* IMG/M 4 version of the integrated metagenome comparative analysis system. *Nucleic Acids Res.* **42,** D568–D573 (2014).
12. Edwards, R. A. & Rohwer, F. Viral metagenomics. *Nat. Rev. Microbiol.* **3,** 504–510 (2005).
13. Ivanova, N. *et al.* A call for standardized classification of metagenome projects. *Environ. Microbiol.* **12,** 1803–1805 (2010).
14. Hurwitz, B. L. U'Ren, J. M. & Youens-Clark, K. Computational prospecting the great viral unknown. *FEMS Microbiol. Lett.* (2016).
15. Ignacio-Espinoza, J. C., Solonenko, S. A. & Sullivan, M. B. The global virome: not as big as we thought? *Curr. Opin. Virol.* **3,** 566–571 (2013).
16. Lu, H. *et al.* Membrane biofouling in a wastewater nitrification reactor: Microbial succession from autotrophic colonization to heterotrophic domination. *Water Res.* **88,** 337–345 (2016).
17. Serwer, P., Hayes, S. J., Thomas, J. A. & Hardies, S. C. Propagating the missing bacteriophages: a large bacteriophage in a new class. *Virol. J.* **4,** 21 (2007).
18. Varghese, N. J. *et al.* Microbial species delineation using whole genome sequences. *Nucleic Acids Res.* **43,** 6761–6771 (2015).
19. Simmonds, P. Methods for virus classification and the challenge of incorporating metagenomic sequence data. *J. Gen. Virol.* **96,** 1193–1206 (2015).
20. Hurwitz, B. L., Brum, J. R. & Sullivan, M. B. Depth-stratified functional and taxonomic niche specialization in the 'core' and 'flexible' Pacific Ocean Virome. *ISME J.* **9,** 472–484 (2015).
21. Roux, S., Hallam, S. J., Woyke, T. & Sullivan, M. B. Viral dark matter and virus-host interactions resolved from publicly available microbial genomes. *eLife* **4,** (2015).
22. Mojica, F. J., Díez-Villaseñor, C., García-Martínez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155,** 733–740 (2009).
23. Andersson, A. F. & Banfield, J. F. Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* **320,** 1047–1050 (2008).
24. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315,** 1709–1712 (2007).
25. Lum, A. G. *et al.* Global transcription of CRISPR loci in the human oral cavity. *BMC Genomics* **16,** 401 (2015).
26. Bailly-Bechet, M., Vergassola, M. & Rocha, E. Causes for the intriguing presence of tRNAs in phages. *Genome Res.* **17,** 1486–1495 (2007).
27. Goren, M. G., Yosef, I. & Qimron, U. Programming Bacteriophages by Swapping Their Specificity Determinants. *Trends Microbiol.* **23,** 744–746 (2015).
28. Salmond, G. P. & Fineran, P. C. A century of the phage: past, present and future. *Nat. Rev. Microbiol.* **13,** 777–786 (2015).
29. Holmfeldt, K., Middelboe, M., Nybroe, O. & Riemann, L. Large variabilities in host strain susceptibility and phage host range govern interactions between lytic marine phages and their Flavobacterium hosts. *Appl. Environ. Microbiol.* **73,** 6730–6739 (2007).
30. Peters, D. L., Lynch, K. H., Stothard, P. & Dennis, J. J. The isolation and characterization of two *Stenotrophomonas maltophilia* bacteriophages capable of cross-taxonomic order infectivity. *BMC Genomics* **16,** 664 (2015).
31. Emerson, J. B. *et al.* Virus-host and CRISPR dynamics in archaea-dominated hypersaline Lake Tyrrell, Victoria, Australia. *Archaea* **2013,** 370871 (2013).
32. Tschitschko, B. *et al.* Antarctic archaea-virus interactions: metaproteome-led analysis of invasion, evasion and adaptation. *ISME J.* **9,** 2094–2107 (2015).
33. Breitbart, M. & Rohwer, F. Here a virus, there a virus, everywhere the same virus? *Trends Microbiol.* **13,** 278–284 (2005).
34. Dinsdale, E. A. *et al.* Functional metagenomic profiling of nine biomes. *Nature* **452,** 629–632 (2008).
35. Breitbart, M., Miyake, J. H. & Rohwer, F. Global distribution of nearly identical phage-encoded DNA sequences. *FEMS Microbiol. Lett.* **236,** 249–256 (2004).
36. Salazar, G. *et al.* Global diversity and biogeography of deep-sea pelagic prokaryotes. *ISME J.* **10,** 596–608 (2016).
37. Abeles, S. R. & Pride, D. T. Molecular bases and role of viruses in the human microbiome. *J. Mol. Biol.* **426,** 3892–3906 (2014).
38. Wylie, K. M. *et al.* Metagenomic analysis of double-stranded DNA viruses in healthy adults. *BMC Biol.* **12,** 71 (2014).
39. Robles-Sikisaka, R. *et al.* Association between living environment and human oral viral ecology. *ISME J.* **7,** 1710–1724 (2013).
40. Mukherjee, S., Huntemann, M., Ivanova, N., Kyrpides, N. C. & Pati, A. Large-scale contamination of microbial isolate genomes by Illumina PhiX control. *Stand. Genomic Sci.* **10,** 18 (2015).
41. Bondy-Denomy, J. & Davidson, A. R. When a virus is not a parasite: the beneficial effects of prophages on bacterial fitness. *J. Microbiol.* **52,** 235–242 (2014).
42. Short, C. M. & Suttle, C. A. Nearly identical bacteriophage structural gene sequences are widely distributed in both marine and freshwater environments. *Appl. Environ. Microbiol.* **71,** 480–486 (2005).
43. Kyrpides, N. C., Eloe-Fadrosh, E. A. & Ivanova, N. N. Microbiome data science: understanding our microbial planet. *Trends Microbiol.* **24,** 425–427 (2016).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Visel and H. Maughan for critical reading and feedback, A. Pati for help in earlier versions, and the IMG and GOLD teams for their support. This work was conducted by the US Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, under contract number DE-AC02-05CH11231 and used resources of the National Energy Research Scientific Computing Center, supported by the Office of Science of the US Department of Energy.

Author Contributions D.P.E., N.N.I., and N.C.K. conceived and led the study. All authors participated in the analysis and interpretation of data. D.P.E., E.E.F., E.R., N.N.I., and N.C.K. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.C.K. (nckyrpides@lbl.gov).

Reviewer Information *Nature* thanks C. A. Suttle and the other anonymous reviewer(s) for their contribution to the peer review of this work.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

Data availability. All the sequence data and metadata from the samples used in this work could be accessed through the Integrated Microbial Genomes with Microbiomes system IMG/M database[11] (https://img.jgi.doe.gov) using both metagenome and scaffold identifiers provided throughout the manuscript and the Supplementary Information. Thus, by using these identifiers in the Genome Search tool or Scaffold Search tool (under 'Find Genomes' tab) in the user interface, the corresponding sequences, their annotations, as well as their associated metadata can be retrieved. Moreover, Hidden Markov Models (HMMs) of viral protein families as well as the predicted DNA viral sequences in fasta format are available at the following public FTP site: (http://portal.nersc.gov/dna/microbial/prokpubs/EarthVirome_DP/).

Metagenomic samples used in this study. All publicly available metagenomic data sets from the IMG/M system (3,042 samples comprising 5 terabase pairs of sequences) were used for this analysis[11]. The sample collection included 1,729 environmental samples, 1,079 host-associated samples, and 234 engineered samples according to the Genomes OnLine sample classification[5]. We identified putative viral contigs in 1,882 out of the 3,042 metagenomic data sets. The metadata for these data sets including sample collection information, library construction and sequencing protocols, as well as assembly strategy were retrieved from GOLD database[5]. Based on GOLD metadata, the vast majority of these data sets were generated from dsDNA using an untargeted approach (that is, only 59 samples underwent viral particle enrichment, viral DNA enrichment or library construction with sequencing protocols optimized for the recovery of viral sequences). All of the data sets have been annotated by the IMG metagenome annotation pipeline[44], which performs gene prediction and functional annotation through assignment of predicted proteins to protein families, such as Pfam[45] and KEGG Orthology (KO) clusters[46]. Some of the data sets included both assembled and unassembled data, while others had only assembled sequences (Supplementary Table 8). An assembly pipeline used for each data set is described in GOLD. In addition, the contiguity of assembled sequences varied greatly from sample to sample. The ecosystem subcategories here used were manually curated according to sample metadata establishing 10 distinct habitat types: marine, freshwater, non-marine-saline and alkaline, thermal springs, terrestrial soil, terrestrial others (including mostly deep subsurface samples), host-associated human, host-associated plants, host-associated others (including host animal-associated other than human), and engineered (for example, bioreactor) (Supplementary Table 8). Only contigs longer than 5 kb (59.5 Gb from 5.1 million contigs) were primarily included in this study.

Normalization factors. We normalized the data sets by the size of the sample (measured as total number of bp from sequences larger than 5 kb) per habitat type. The normalization factor used in each habitat type was: marine, 29,602 Mb; freshwater, 96,314 Mb; non-marine saline and alkaline, 2,825 Mb; thermal springs, 1,828 Mb; terrestrial (soil), 5,794 Mb; terrestrial (other), 1,659 Mb; host-associated (human), 10,349 Mb; host-associated (plants), 3,909 Mb; host-associated (others), 23,452 Mb; engineered, 10,486 Mb.

Isolate reference viruses (iVGs). We used a combination of 2,353 iVGs composed of all isolate dsDNA viruses and retroviruses from the NCBI server (http://www.ncbi.nlm.nih.gov/genome/viruses/, data accessed on 04/2015) to extract all viral proteins and to establish, after filtering, the first round of viral protein families. Additionally, we used a list of 5,042 reference viruses (Supplementary Table 13) extracted from the IMG/M system (that included all RNA and DNA eukaryotic and prokaryotic referenced viral genomes) to generate and validate our viral genome clustering method and also to calculate the average length of all reference viruses as 44,296 bp $\pm$ 83,777 bp s.d. (Supplementary Table 13).

Generation of viral protein families. 167,042 protein coding genes were collected from 2,353 iVGs (dsDNA viruses and retroviruses combined) from the NCBI server (http://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=10239#). After dereplication using 70% identity in usearch[47], 98,000 protein sequences were obtained from which 83,500 were clustered into 15,900 groups using the Markov Cluster (MCL) algorithm[48]. Proteins within clusters were aligned using MAFFT[49] and a set of 14,296 viral protein families was created using hmmbuild[50]. After manual curation of the viral families with high representation in prokaryotic genomes, viral protein families were compared against the 5.1 million metagenomic contigs longer than 5 kb. 62,000 contigs with 5 or more viral protein families were collected, and these were reduced to 9,000 putative viral contigs after removing contigs below 50 kb. An additional filtering step was performed to exclude contigs with a high number of Kegg Orthology (KO) terms and Pfams (10% and 25% respectively); this reduced the number of putative viral contigs to 1,589. These were complemented with 66 and 188 sequences derived from diverse metagenomic contigs longer than 20 kb that were binned with viruses or contained a viral RNA polymerase gene, respectively, and were not captured using the previous filter of bearing 5 or more viral protein families (detailed in section below; Extended Data Fig. 2; Supplementary Table 1). A total of 1,843 mVCs encoding 191,000 proteins were used to complement the original set of 167,042 proteins derived from iVGs. Repeating the steps described above (that is, *usearch* 70% for de-replication, MCL clustering, MAFFT alignment and *hmmbuild* with a filter for viral families abundant in prokaryotes) the final list of 25,000 viral protein families was obtained and used for further exploration.

Identification of metagenomic viral contigs for a training set via manual curation, binning and DNA-dependent RNA polymerase alignment. To expand the training set of viral sequences, metagenome contigs identified as high-confidence viral sequences in the first iteration of our pipeline (Extended Data Fig. 1) were complemented with additional metagenome contigs and scaffolds, not captured using viral protein families generated from isolate viruses. The first approach used kmer-based binning of 6 metagenome samples that contained the highest number of candidate viral sequences, which were not satisfy high-confidence threshold due to insufficient number of hits to protein models. These data sets were binned by Emergent Self Organizing Maps (ESOM; by Ultsch) as described previously[51] and contig sets outside the bins corresponding to cellular organisms were manually checked (Extended Data Fig. 2a). K-mer-based binning identified 66 putative novel mVCs from diverse habitat types (freshwater, wastewater, thermal vents, and marine with IMG sample identifiers 3300000553, 3300001592, 3300001681, 3300000116, and 3300001450, respectively).

The second approach relied on identification of contigs containing RNA polymerase with domain composition reminiscent of RNA polymerase (RNAp) found in cellular life forms, which could not be placed into one of three domains on the tree of life based on their sequence similarity. First, 2,551 representative sequences of the genes encoding the three major subunits (α, β, β') of the RNAp gene from bacteria, as well as their eukaryotic and archaeal counterparts, were collected from IMG database. Next, the domains of these genes were extracted using Pfam models and aligned with MAFFT[49]. Alignments were manually inspected and HMM models were built using hmmbuild[50]. These models were used to scan metagenomic sequences longer than 5 kb and identified 39,109 contigs with matches for at least one core RNAp subunit. After filtering short matches and a dereplication step, we obtained 7,437 metagenomic sequences that were combined with 2,551 reference isolates to build a tree with 9,309 RNAp sequences using FastTree[52] with default parameters (Extended Data Fig. 2d). The tree was visualized using Dendroscope[53] and RNAp branch corresponding to large eukaryotic DNA viruses was identified on the basis of reference sequences from isolate genomes. In addition to eukaryotic viruses, another set of metagenomic RNAp sequences branching separately from cellular references, turned out to comprise phage RNAp with domain composition similar to bacterial enzyme (detailed in Extended Data Fig. 2e). A total of 188 contigs longer than 20 kb containing viral and phage RNAp sequences were added to the training set.

Assignment of metagenomic sequences to viruses. The 25,000 viral protein families were used to identify 125,842 DNA metagenomic viral contigs (mVCs) longer than 5 kb using 3 distinct filters. First, mVCs were identified from metagenomic contigs that had at least 5 hits to viral protein families, the total number of genes covered with KO terms on the contig was $<$20%; the total number of genes covered with Pfams $\leq$40%; and the number of genes covered with viral protein families $>$10%. Second, metagenomic sequences were selected as mVCs when the number of viral protein families on the contig were equal or higher than the number of Pfams. Finally, metagenomic contigs for which the number of viral protein families was equal or higher than 60% of the total of the genes were also assigned to mVCs. Benchmarking and modelling of this DNA viral discovery computational approach are detailed below, demonstrating a specificity of 99.6% for viral detection with a 37.5% recall rate (sensitivity to identify all viral sequences).

Benchmarking of computational approaches for virus detection. In order to assess the accuracy of our DNA vHMM virus detection pipeline, we generated a synthetic metagenome, consisting of finished genomes of 32 bacteria, 3 archaea and 5 viruses (Supplementary Table 2), which included 88 replicons. Bacterial genomes include representatives of 4 phyla. A total of 132 prophage sequences were identified including 99 prophages identified by CyVerse[54] implementation of VirSorter[55] in the categories 1, 2, 4, and 5, and 33 prophages identified by manual curation based on the presence of hallmark phage genes and analysis of synteny with closely related strains. Coordinates of 35 prophages predicted by VirSorter had to be manually adjusted to eliminate bacterial genes (including ribosomal RNAs and other housekeeping genes) and to separate 2 prophage sequences called as one prophage over an intervening stretch of bacterial genes. Coordinates of the prophages are provided in Supplementary Table 3. None of the viruses or prophages used in the synthetic metagenome were included in the training set used to generate viral HMMs for our pipeline.

The genome sequences were fragmented to generate 63,222 contigs of length 5 kb to 60 kb. The distribution of fragmented contigs include 28,497 5-kb-long

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

fragments, 14,228 10-kb-long fragments, 7,096 20-kb-long fragments, 4,723 30-kb-long fragments, 3,525 40-kb-long fragments, 2,810 50-kb-long fragments and 2,343 60-kb-long fragments. The resulting synthetic metagenome is dominated by bacterial and archaeal chromosomal fragments with an admixture of a relatively small number of plasmid and viral sequences, which is a faithful representation of a typical metagenome data set generated by an untargeted approach rather than by targeted virome sequencing approach. The metagenome was submitted to a CyVerse implementation of VirSorter and also processed by our vHMM pipeline. Only the categories 1, 2, 4 and 5 of Virsorter predictions were considered, as manual inspection showed that categories 3 and 6 contained mostly false positives. Sequence fragments with at least 3 kb of phage or prophage sequence were considered as true positive viral sequences; those with less than 3 kb of phage or prophage sequence were considered true negative.

Calculating the rate of viral protein cluster accumulation and the number of proteins with high similarity to proteins encoded by isolate viruses. The 125,842 metagenomic viral contigs longer than 5 kb encoded a total of 2.79 million proteins. BLASTp[56] with an *e*-value of 1.0×10^{-5} was used and 1 hit per query protein with >60% sequence identity and >80% alignment on the shorter sequence. Proteins encoded by mVCs were clustered using CD-HIT[57] at 60% sequence identity and >80% alignment on the shorter sequence. For each sample count, 100 random metagenome sets were generated and the total number of protein clusters found on the contigs from this set was calculated. This analysis was repeated separately for metagenome samples classified as 'aquatic' ($n = 656$) and 'human' ($n = 673$).

Comparison of mVCs protein clusters against all iVGs. Sequence similarity of mVCs to iVGs was computed using BLASTp[56] with an *e*-value threshold of 1.0×10^{-5} and alignment length of at least 80% of the shorter protein. No percentage identity or bit-score cutoffs were applied (Supplementary Table 9).

Identification of complete metagenomic viral genomes. To assess the number of closed DNA mVCs, we searched for overlapping sequences in the 3′ and 5′ region of all the 125,842 metagenomic contigs. Extractseq[58] was used to trim the first 100 bp of each contig and BLAT[59] was used to search each 100-bp fragment against the respective contig. Only exact overlapping matches for both the 3′ and 5′ regions were considered. This resulted in the identification of 999 putatively closed mVCs, ranging from 5,037 bp to 630,638 bp in length (average, 53,644 bp ± 45,677 bp s.d.). Supplementary Table 10 lists all putatively closed mVCs.

Viral genome clustering and designation of viral groups. A sequence-based classification framework was developed for systematically linking closely related viral genomes based on their overall protein similarity. The framework relies on both AAI and total alignment fraction (AF) for pairwise comparisons of viral sequences, and enables natural grouping of related iVGs and mVCs. The 125,842 mVCs were combined with all iVGs (DNA and RNA viruses) for the generation of the viral group classification framework (Supplementary Information). To reduce the number of the AAI comparisons, only mVCs that contained at least one protein match with ≥70% identity across ≥50% of the shortest protein length were selected for pairwise computations. This filter reduced the number of total pairwise comparisons from 9.5 billion to 15.9 million. The bidirectional average amino acid identity (AAI) was performed as previously described[18] for all of the 15.9 million pairwise comparisons. This method implements usearch[47] for rapid blast, and selects the bidirectional best hit for each protein encoded on the mVC and outputs the AAI and the AF. The output was subsequently filtered to include only matches that had ≥90% AAI and ≥50% AF which were the observed parameters that best reproduced the existing taxonomy of iVGs (Supplementary Information; Extended Data Fig. 5a). The high-quality filtered AAI results were then clustered using single-linkage hierarchical clustering and visualized in Cytoscape[60] (Extended Data Fig. 5c–e).

Validation of viral groups generated. As a validation of our clustering method we observed that 87% of the iVGs (920 out of the 1,060 viral groups or singletons) with a taxonomic assignment according to the International Committee on Taxonomy of Viruses (ICTV) clustered in agreement with the ICTV-designated species. All the remaining 13% of iVGs clustered at the genus-level. From this 13% (represented by 140 viral groups that contain at least one iVG) we found that only 49 were phage groups, with high pairwise (over 90% AAI) values for the reference viruses within each group, suggesting that despite their taxonomic assignments, they were also probably members of the same species (Supplementary Table 14). These analyses show that our viral groups are taxonomically relevant and provide a useful method for organizing distinct viral types.

Viral host assignment using the CRISPR–Cas system. A CRISPR–Cas spacer database of 3.5 million sequences was created using a modified version of the CRISPR Recognition Tool[61] (CRT) detailed in ref. 44 against 40,623 isolates and 6,714 metagenomes (all data sets from the IMG system as of 9 July 2015). All identified spacers were queried for exact sequence matches against all iVGs using the BLASTn-short function from the BLAST+ package with parameters: *e*-value threshold of 1.0×10^{-10}, percentage identity of 95%, and using 1 as a maximum target sequence[56]. 98.5% of the detected 1,340 spacer hits were to a putative bacterial or archaeal host whose taxonomic assignment was in agreement at the species or genus level with the existing viral taxonomy (Supplementary Table 18). From the remaining matches, 1.2% of the hits agreed at the family level and only 0.3% of the spacers (2 cases where *Pseudomonas* spacers matched a *Rhodothermus* phage, and *Methylomicrobium* spacers that matched *Pseudomonas* and *Burkholderia* phage) were above family, validating our approach of host assignment based on CRISPR–Cas spacer matches. Subsequently, all 3.5 million spacers were compared against the 125,842 mVCs, requiring at least 95% identity over the whole spacer length, and allowing only 1–2 SNPs at the 5′ end of the sequence. A total of 12,576 proto-spacers (that is, spacer sequences within a phage genome) were identified. Based on CRISPR–Cas spacer matches exclusively from microbial isolate genomes we assigned host taxonomy to 8,084 mVCs (representing 6.42% of all the mVCs), comprising 826 viral groups (~4.47% of the total) plus 1,100 viral singletons (~1.71%) (Fig. 2a; Supplementary Table 19).

Host–virus assignment using viral tRNA matches. Identification of tRNAs from mVCs was performed with ARAGORN v1.2 (ref. 62) using the '-t' option. In order to validate this approach, 2,181 tRNA sequences were recovered from 344 referenced viruses (~7% of the total). These were compared against all genomes and metagenomes in the IMG system using BLAST, leading to 16,089 perfect hits (100% length and 100% sequence identity) after removing self-hits and duplicates. The taxonomic assignment of the tRNAs found in iVGs was compared against the taxonomic information of the isolate microbial genomes showing that 92.5% of the matches agreed at the genus or species level (Supplementary Table 18). After culling the top-20 most abundant viral-tRNA sequences (sequences conserved across members of the gammaproteobacteria class; Supplementary Table 22) and repeating the above steps with mVCs, 32,449 tRNAs within 9,555 mVCs (7.6% out of the 125,842 total) were identified, enabling the host assignment for 2,527 mVCs (Supplementary Information; Supplementary Table 19).

Low abundance virus detection. In order to detect the presence of any of the mVCs in lower abundances across different habitat types, we expanded our analysis to include not only assembled data (that probably represent the most abundant viruses) but also unassembled data from 4,169 samples currently available in IMG/M database, which comprises more than 5 Tb of sequences. We used BLASTn program in the Blast+ package[56] to find hits to our 125,842 predicted viral sequences with an *e*-value cutoff of 1×10^{-50}, at least 90% identity, and the hits from the sample covering at least 10% of the length of the viral contig. This filtering of BLAST results excluded matches to short highly conserved fragments of viral sequences, such as tRNAs, and other spurious hits. Our filtering criteria were optimized for the type of metagenome data sets available to us, and are significantly more stringent than those used in some previous studies for similar data (e.g. 95% identity over 75 nt alignment used in ref. 63) or tBLASTx with *e*-value of 1.0×10^{-5} recommended by ref. 64. However, it was less stringent than the 75% coverage used in the analysis of Tara Oceans Viromes[3], which relied on viral enrichment to increase viral sequence coverage. For the largest metagenome available to us (IMG taxon 3300002568, Grasslands soil microbial communities from Hopland, California, USA), this new analysis was able to detect 500 nt of viral sequence in 138,769,704,035 nt of total metagenome sequence, which corresponds to the abundance of 3.06×10^{-07}%.

Habitat type specificity of predicted viral sequences based on their BLASTn hits in assembled and unassembled data shows their presence even at low abundance, depending on the sequence coverage for each specific metagenome (Fig. 4a, b). The distribution of less abundant viruses supports the trend that viruses have a strong specificity for a particular habitat type since ~84% of all the viral groups are found exclusively in a single habitat type. About 14% of the viral groups were found in 2 habitat types, and most of these cases could be explained by the uncertainty of habitat type classification. For instance, algae-associated microbiomes were classified as plant host-associated and shared viral groups with marine samples, whereas loose soil samples classified as terrestrial habitat type shared viral groups with rhizosphere samples, which were classified as plant host-associated. After excluding ambiguously classified cases, most viral groups detected in more than 1 habitat type were found in the samples from the same environmental category (for example, in different aquatic habitats or in different mammalian hosts). We further report the finding of ~0.2% of the viral groups in 5 or more habitats types and discuss the main types of these 'cosmopolitan' viruses (probably laboratory contaminants, prophages with broad-host specificity, and bona fide lytic phages with unexpectedly broad habitat type distribution).

Estimation of viral abundance in faecal and oral metagenomes from Human Microbiome Project. Raw reads for faecal and oral metagenomes were retrieved from the Short Read Archive (http://www.ncbi.nlm.nih.gov/sra/) based on the metadata available in GOLD. The reads were quality-filtered and quality-trimmed

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

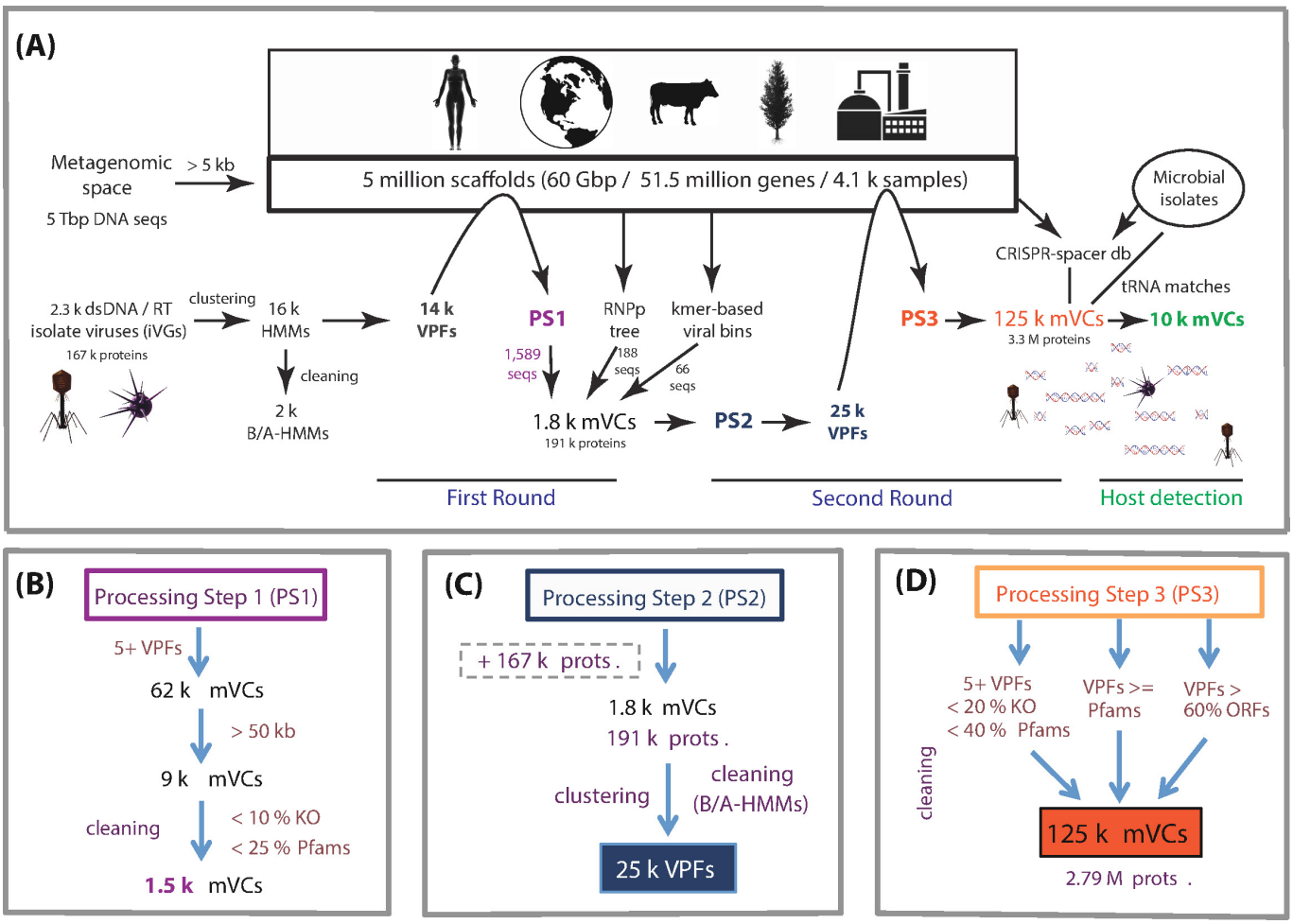
using rqcfilter tool from BBtools package (https://sourceforge.net/projects/bbtools/) with default settings: kmer length for trimming of 23, minimum average quality of 5, trim quality threshold of 10, reads shorter than 45 nt after trimming were discarded. Quality-filtered and trimmed reads were digitally normalized and error corrected using bbnorm tool from BBtools package with default settings. Normalized reads were assembled using SPADES 3.6.2 (ref. 65) and kmers of 19, 39, 59, 79, 99, selecting an optimal kmer length based on the maximal N50. Average contig and scaffold coverage of assembled data was calculated by mapping the quality-filtered and -trimmed reads to the assembly using bbmap tool from BBtools with default kmer length of 13 and minimum percentage identity cutoff of 95%. The unmapped reads were merged using bbmerge tool from BBtools package and the sequences shorter than 100 nt were discarded. mVCs were aligned against these data using BLASTn and filtered as described above. Only 1 best hit per sequence was retained. Coverage of each mVC by sample data was calculated as alignment length multiplied by the coverage of the subject sequence and summed over all sequences in the sample with hits to this mVC.

Putative prophage identification. We have identified putative prophages among 125,842 mVCs using these contigs as a query and running BLASTn[56] comparison of 'blast+' package against all isolate genomes in the IMG database. *e*-value cutoff of 1.0×10^{-50} and percentage identity of 80% were used, and mVCs with cumulative alignment of at least 75% of mVC length against an isolate genome were considered prophage candidates (Supplementary Table 4).

Global virus distribution maps. Visualization was made with the use of Processing programming language (https://processing.org/) and a freely available equirectangular projection of the world map (http://eoimages.gsfc.nasa.gov/images/imagerecords/57000/57752/land_shallow_topo_2048.jpg) was used as a background image. Sample points are positioned by latitude and longitude coordinates of Biosamples obtained from GOLD[5]. Points are coloured based on a customized reclassification of the GOLD hierarchical ecosystem classification (habitat types). Lines between points indicate samples that share at least 2 viral groups or singletons.

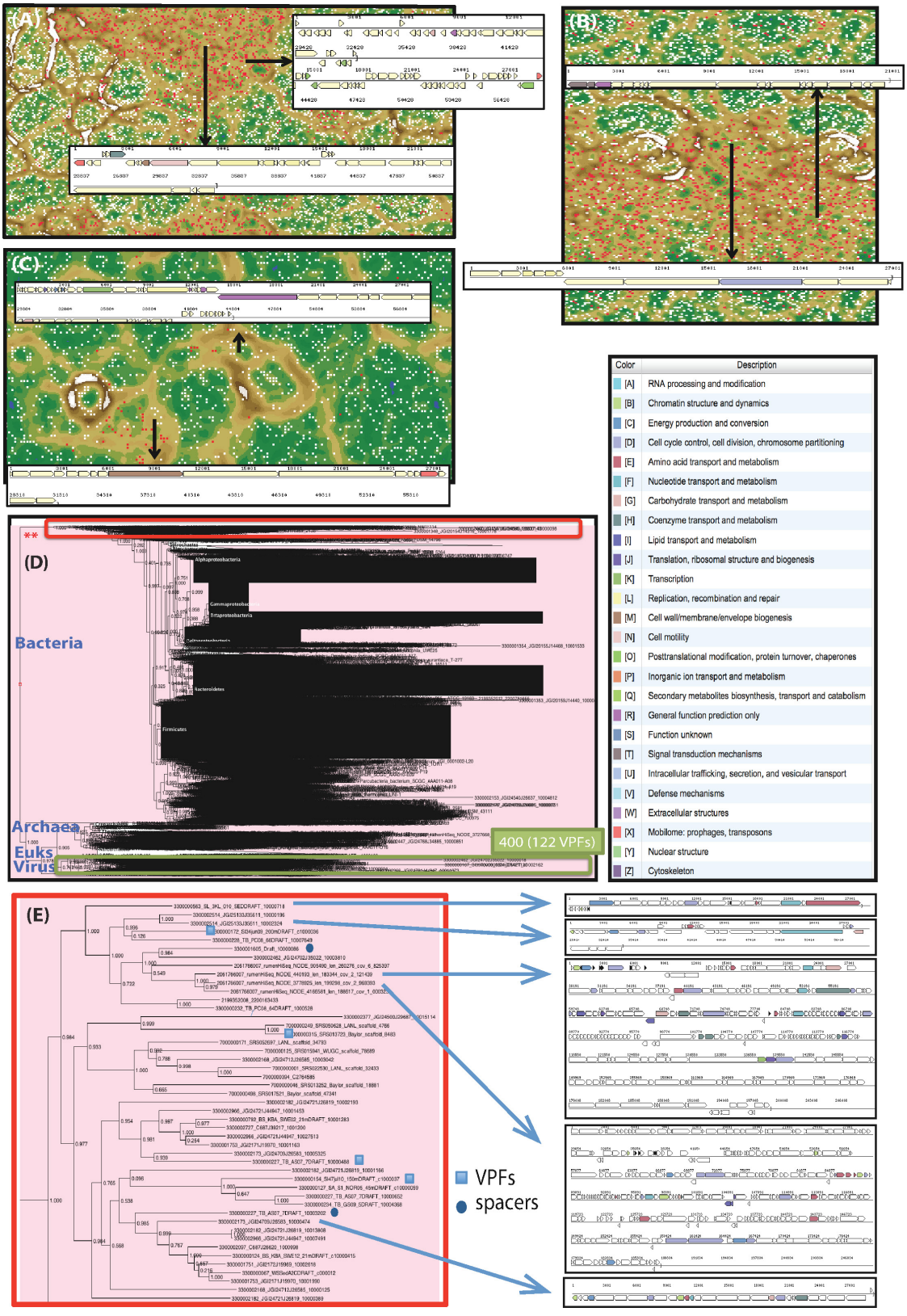
44. Huntemann, M. *et al.* The standard operating procedure of the DOE-JGI Microbial Genome Annotation Pipeline (MGAP v.4). *Stand. Genomic Sci.* **10,** 86 (2015).
45. Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44** (D1), D279–D285 (2016).
46. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44** (D1), D457–D462 (2016).
47. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26,** 2460–2461 (2010).
48. Enright, A. J., Van Dongen, S. & Ouzounis, C. A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30,** 1575–1584 (2002).
49. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30,** 3059–3066 (2002).
50. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* **39,** W29–W37 (2011).
51. Dick, G. J. *et al.* Community-wide analysis of microbial genome sequence signatures. *Genome Biol.* **10,** R85 (2009).
52. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26,** 1641–1650 (2009).
53. Huson, D. H. & Scornavacca, C. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Syst. Biol.* **61,** 1061–1067 (2012).
54. Merchant, N. *et al.* The iPlant Collaborative: Cyberinfrastructure for Enabling Data to Discovery for the Life Sciences. *PLoS Biol.* **14,** e1002342 (2016).
55. Roux, S., Enault, F., Hurwitz, B. L. & Sullivan, M. B. VirSorter: mining viral signal from microbial genomic data. *PeerJ* **3,** e985 (2015).
56. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10,** 421 (2009).
57. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22,** 1658–1659 (2006).
58. Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* **16,** 276–277 (2000).
59. Kent, W. J. BLAT–the BLAST-like alignment tool. *Genome Res.* **12,** 656–664 (2002).
60. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13,** 2498–2504 (2003).
61. Bland, C. *et al.* CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* **8,** 209 (2007).
62. Laslett, D. & Canback, B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* **32,** 11–16 (2004).
63. Dutilh, B. E. *et al.* A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nat. Commun.* **5,** 4498 (2014).
64. Aziz, R. K., Dwivedi, B., Akhter, S., Breitbart, M. & Edwards, R. A. Multidimensional metrics for estimating phage abundance, distribution, gene density, and sequence coverage in metagenomes. *Front. Microbiol.* **6,** 381 (2015).
65. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19,** 455–477 (2012).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 1 | Detailed workflow for the identification of viral sequences from metagenomic data. a, Overview of the acquisition and filtering of viral protein families in two rounds and their use for the identification of metagenomic viral contigs larger than 5 kb. In the first round, proteins from 2,300 double-stranded DNA viruses were grouped into 16,000 protein families, which were aligned to generate Hidden Markov Models (HMMs). These HMMs were used in combination with analysis of k-mer composition and phylogenetic analysis of DNA-dependent RNA polymerase genes to identify 1,843 high-confidence metagenome viral contigs. **b**, **c**, These contigs were validated by manual analysis (**b**) and the proteins from this set were combined with the isolate viral proteins to generate a final set of 25,000 viral protein families (**c**). **d**, HMMs generated from alignment of these protein families were used to identify 125,842 metagenomic viral contigs. Processing steps detailed in **b**–**d** are described in the Methods. The final mVCs were then grouped and assigned to their hosts via CRISPR–Cas spacer matches and viral tRNA matches against isolate microbes (not shown in this figure).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

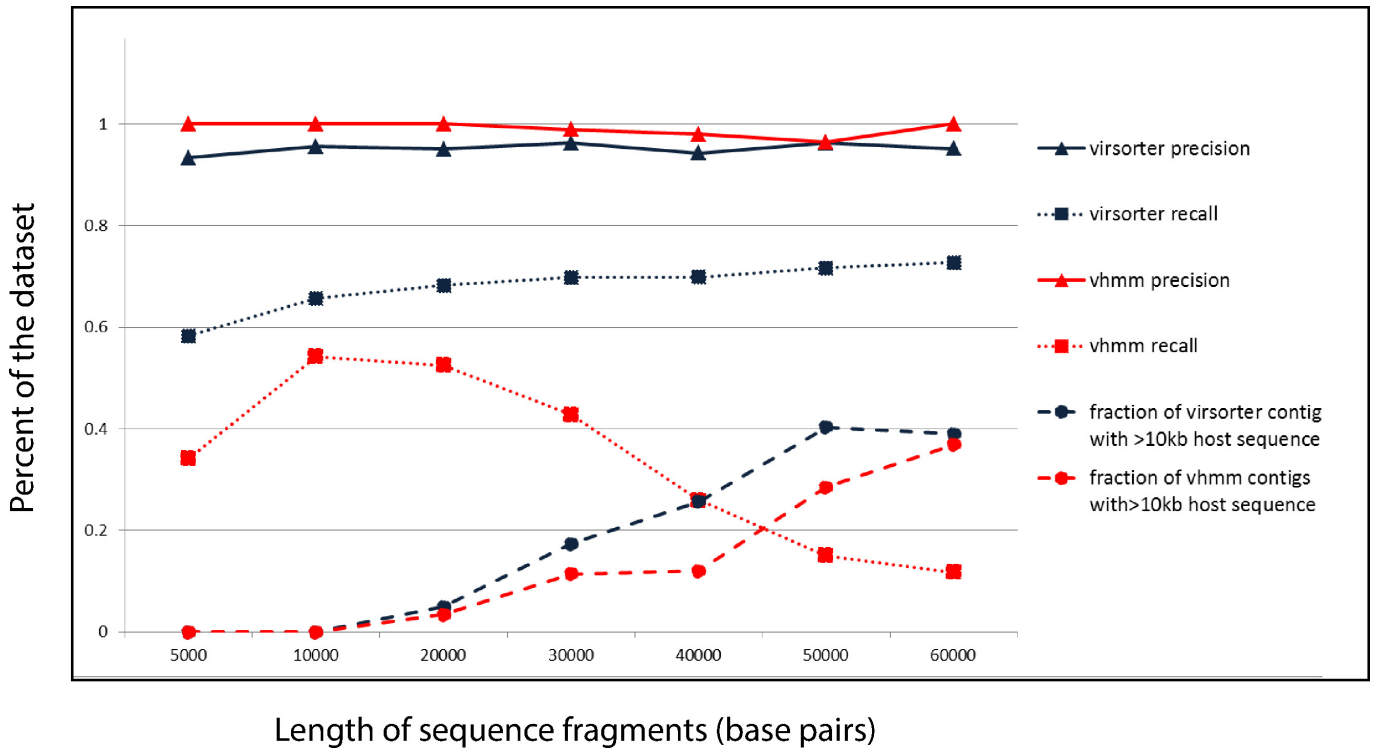


Extended Data Figure 2 | See next page for caption.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

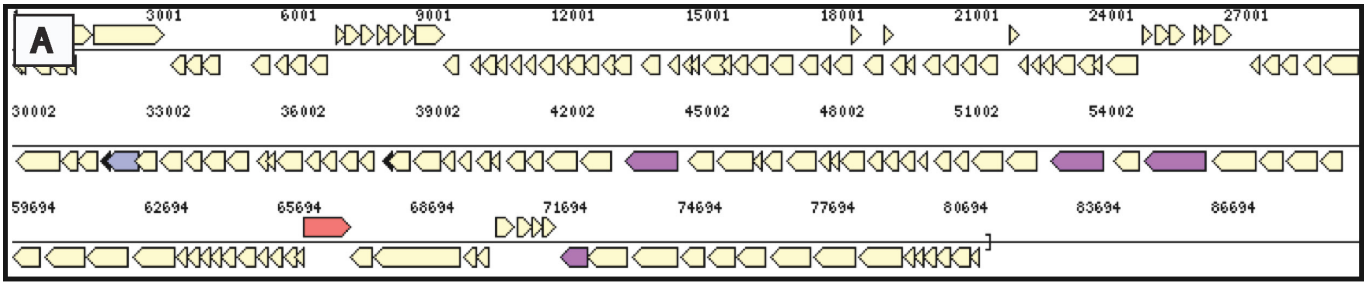
Extended Data Figure 2 | Identification of metagenomic viral contigs via binning and DNA-dependent RNA polymerase alignment. **a**–**c**, Three distinct metagenomic examples of tetranucleotide Emergent Self Organizing Maps (ESOM) as a binning method for identification of candidate viral sequences in metagenome data sets. Tetranucleotide binning of metagenomic samples (full list in Supplementary Table 1) was used to identify highly divergent viral sequences, which were left undetected using viral protein families generated from isolate viruses. Each dot on the maps represents a 10 kb fragment of a metagenomic scaffold longer than 20 kb. 'Bubbles' (ESOM structures) correspond to fragments with similar tetranucleotide composition probably originating from the same genome. Red dots represent viral sequences detected by viral protein families generated for isolate viruses; white dots represent highly divergent viral sequences with no hits to viral protein families. **a**, ESOM of freshwater sample (combined assembly of freshwater microbial communities from Lake Mendota and Trout Bog Lake, IMG identifier 3300000553) shows 2 putative viral sequences previously unidentified (IMG scaffold identifiers 10001161 and 10001271). **b**, ESOM of marine sample (marine microbial communities from Delaware Coast, sample from Delaware MO Spring March 2010, IMG identifier 3300000116) shows 2 putative viral sequences sequences (IMG scaffold identifiers c10000689 and c10000429). **c**, ESOM of hydrothermal vent sample (black smokers hydrothermal plume microbial communities from Abe, Lau Basin, Pacific Ocean, IMG identifier 3300001681) showing 2 viral sequences (IMG scaffold identifiers 10000222 and 10000095). Metagenome samples can be found in IMG using IMG identifiers and 'Quick Search' or 'Genome Search' tools; metagenome scaffolds can be using scaffold identifier and 'Scaffold Search' tool on the respective 'Microbiome Details' page. **d**, **e**, DNA-dependent RNA polymerase genes of likely viral origin from metagenomic sequences longer than 5 kb. **d**, Hidden Markov Models (HMMs) were built for sequences corresponding to α, β, and β' subunits of bacterial DNA-dependent RNA polymerase for a representative set of 2,551 cellular organisms (archaea, bacteria, and eukaryotes) and viruses. These models were used to search the proteins encoded by metagenomic contigs longer than 5 kb and the proteins with hits were aligned against the HMMs. A total of 7,437 nearly full-length metagenomic sequences were combined with 2,551 reference sequences to reconstruct the phylogenetic tree using FastTree tool. Two distinct branches on this tree were separated from the sequences from cellular organisms and included RNA polymerase genes from eukaryotic viruses (green box) and putative phage sequences with domain structure similar to that of bacterial RNA polymerase (red box, marked with double asterisk). Only 122 out of the 400 contigs in the eukaryotic viral RNA polymerase branch were captured by isolate protein families. **e**, Detailed view of the RNA polymerase tree branch with putative phage sequences. Metagenome contigs detected as viral by viral protein families and by spacer hits are marked with a square or circle next to it. Gene structure for selected contigs (IMG chromosomal neighbourhood view) is shown in the boxes. In the examples, genes are coloured based on predicted function category (using Clusters of Orthologous Genes prediction) and are specified in the figure. White-coloured genes correspond to those with hypothetical or unknown function.

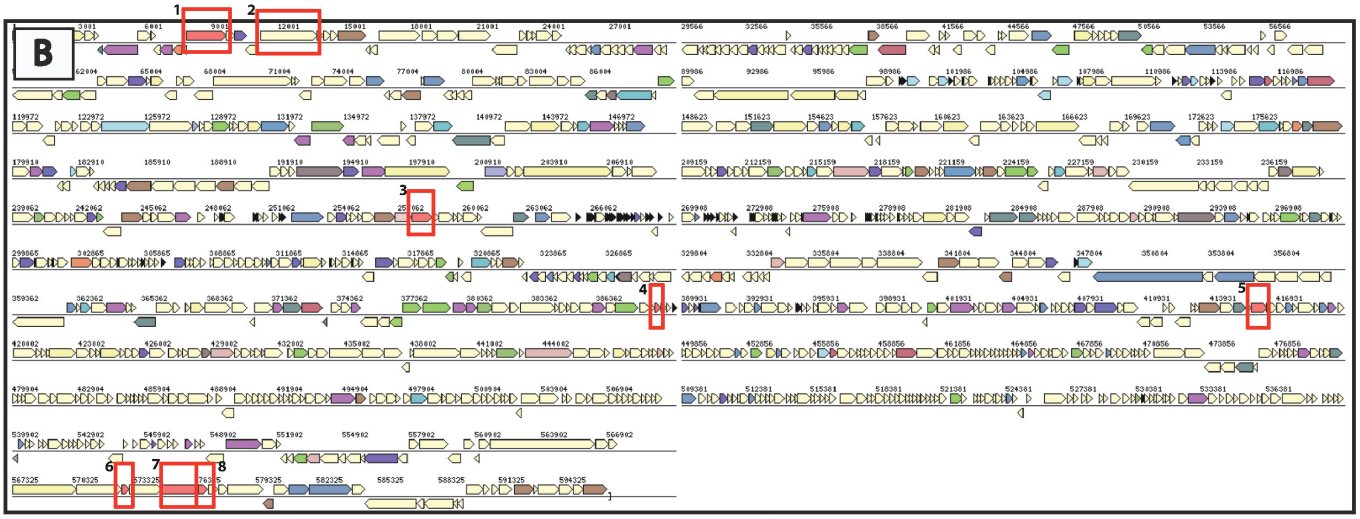
© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 3 | Benchmarking of vHMM-based pipeline and VirSorter on synthetic metagenome data. Precision (solid lines) and recall (dotted lines) for vHMM pipeline (red) and VirSorter (blue) is plotted against the length of sequence fragments in base pairs. The percentage of contigs detected as viral, but which have at least 10 kb of host sequence is shown by dashed lines for vHMM pipeline (red) and VirSorter (blue).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



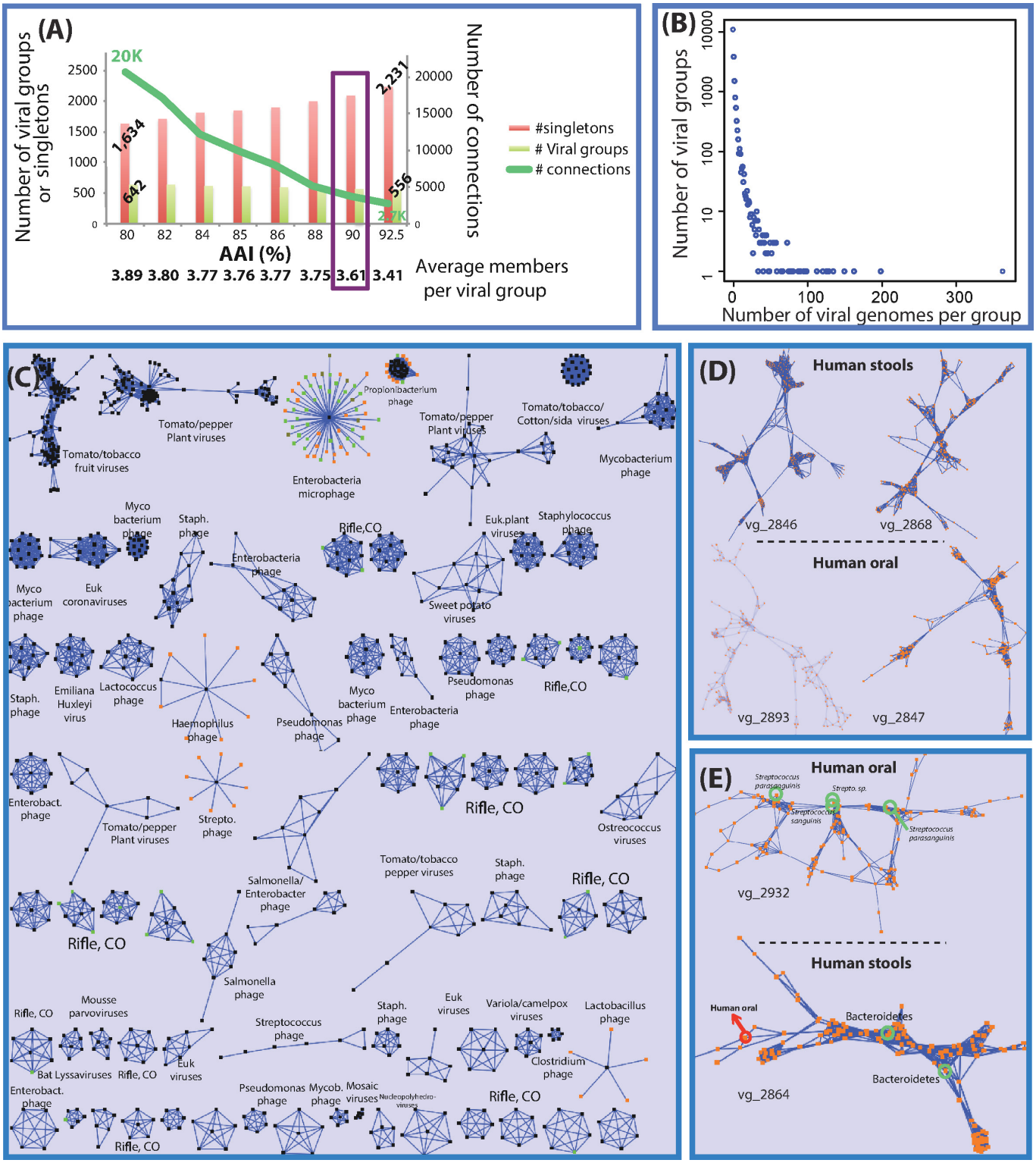


Color	Description
[A]	RNA processing and modification
[B]	Chromatin structure and dynamics
[C]	Energy production and conversion
[D]	Cell cycle control, cell division, chromosome partitioning
[E]	Amino acid transport and metabolism
[F]	Nucleotide transport and metabolism
[G]	Carbohydrate transport and metabolism
[H]	Coenzyme transport and metabolism
[I]	Lipid transport and metabolism
[J]	Translation, ribosomal structure and biogenesis
[K]	Transcription
[L]	Replication, recombination and repair
[M]	Cell wall/membrane/envelope biogenesis
[N]	Cell motility
[O]	Posttranslational modification, protein turnover, chaperones
[P]	Inorganic ion transport and metabolism
[Q]	Secondary metabolites biosynthesis, transport and catabolism
[R]	General function prediction only
[S]	Function unknown
[T]	Signal transduction mechanisms
[U]	Intracellular trafficking, secretion, and vesicular transport
[V]	Defense mechanisms
[W]	Extracellular structures
[X]	Mobilome: prophages, transposons
[Y]	Nuclear structure
[Z]	Cytoskeleton

	Gene product name	Start Coord	End Coord	Strand
1	Terminase	7,846	9,678	+
2	T4-like capsid prot (gp20)	11,212	13,776	+
3	Mu-like tail prot gpP	257,685	258,626	+
4	Phage repressor prot C	389,035	389,328	+
5	Phage prot gp37	415,752	416,477	+
6	Phage baseplate assembly prot W	572,696	573,109	+
7	Phage tail prot (Tail P2 I)	574,656	576,485	+
8	Tail fiber prot	576,482	576,946	+

Extended Data Figure 4 | Detailed gene content of singular metagenomic viral contigs examples. a, Gene content of the metagenomic partial viral genome with the lowest gene coverage by viral protein families. This length of the partial viral genome is 81,542 bp (guanine and cytosine (GC) content of 43%; 163 total genes) and was identified from a bovine rumen metagenome (IMG scaffold identifier, rumenHiSeq_NODE_3763566_len_81492_cov_5_518198; IMG metagenome identifier, 2061766007). White-coloured genes correspond to those with hypothetical or unknown function. Only 3% of the genes were covered by VPFs. **b**, Gene content of the largest closed viral genome identified to date. The length of the closed (circular) viral genome is 596,617 bp (GC, 40%; 1,148 total genes) and was identified from a bioreactor metagenome (IMG scaffold id: D1draft_1000006, from Bioreactor L1-648F-DHS sludge microbial communities sample). Predicted gene function is coloured based on Clusters of Orthologous Genes. Black triangles indicate tRNAs sequences (**a**, **b**). A total of 11% of the genes were covered by VPFs. Specific viral genes distributed across the genome are boxed in red, identified with a number, and described in the legend table. The detailed information of the whole gene content of this viral genome is located in Supplementary Table 11.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

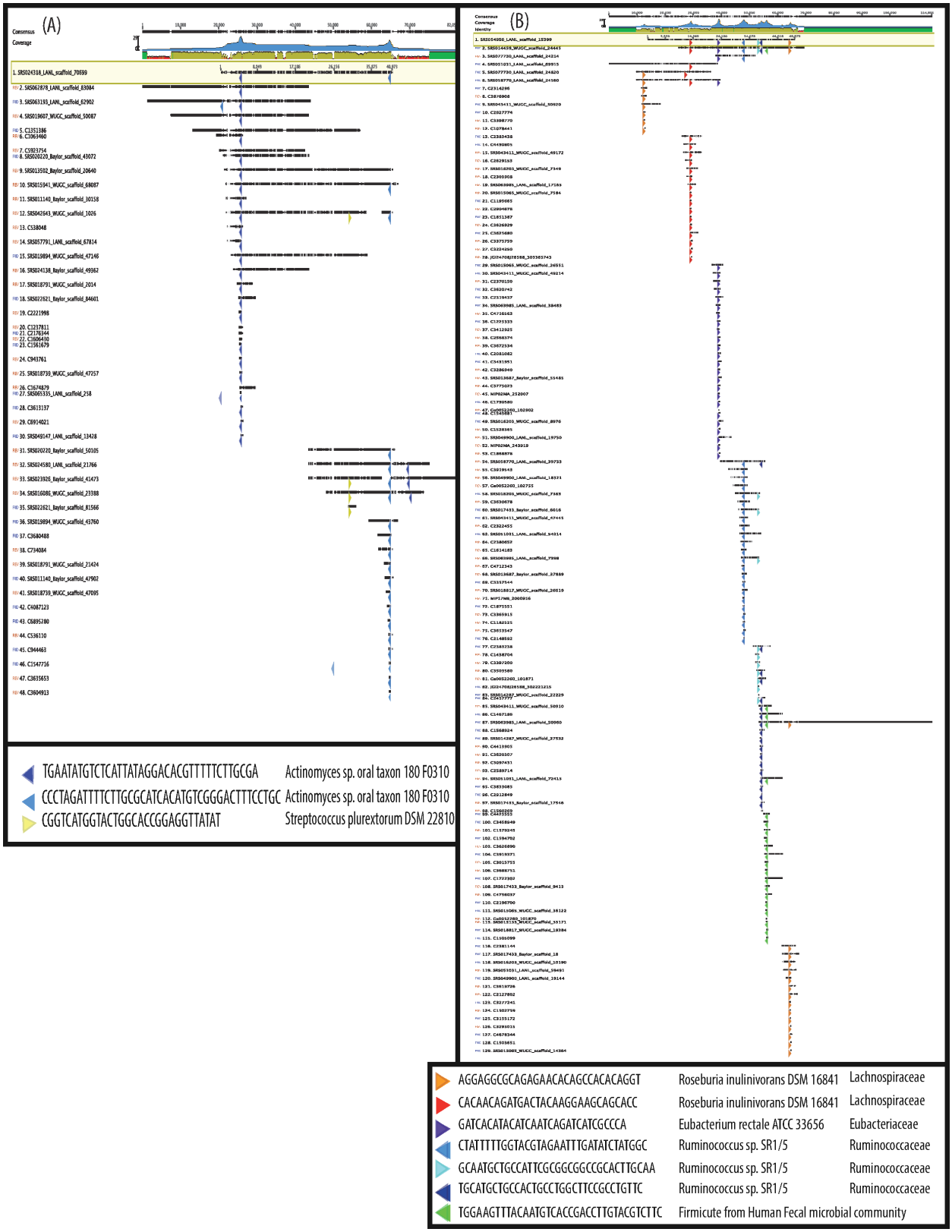


Extended Data Figure 5 | See next page for caption.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

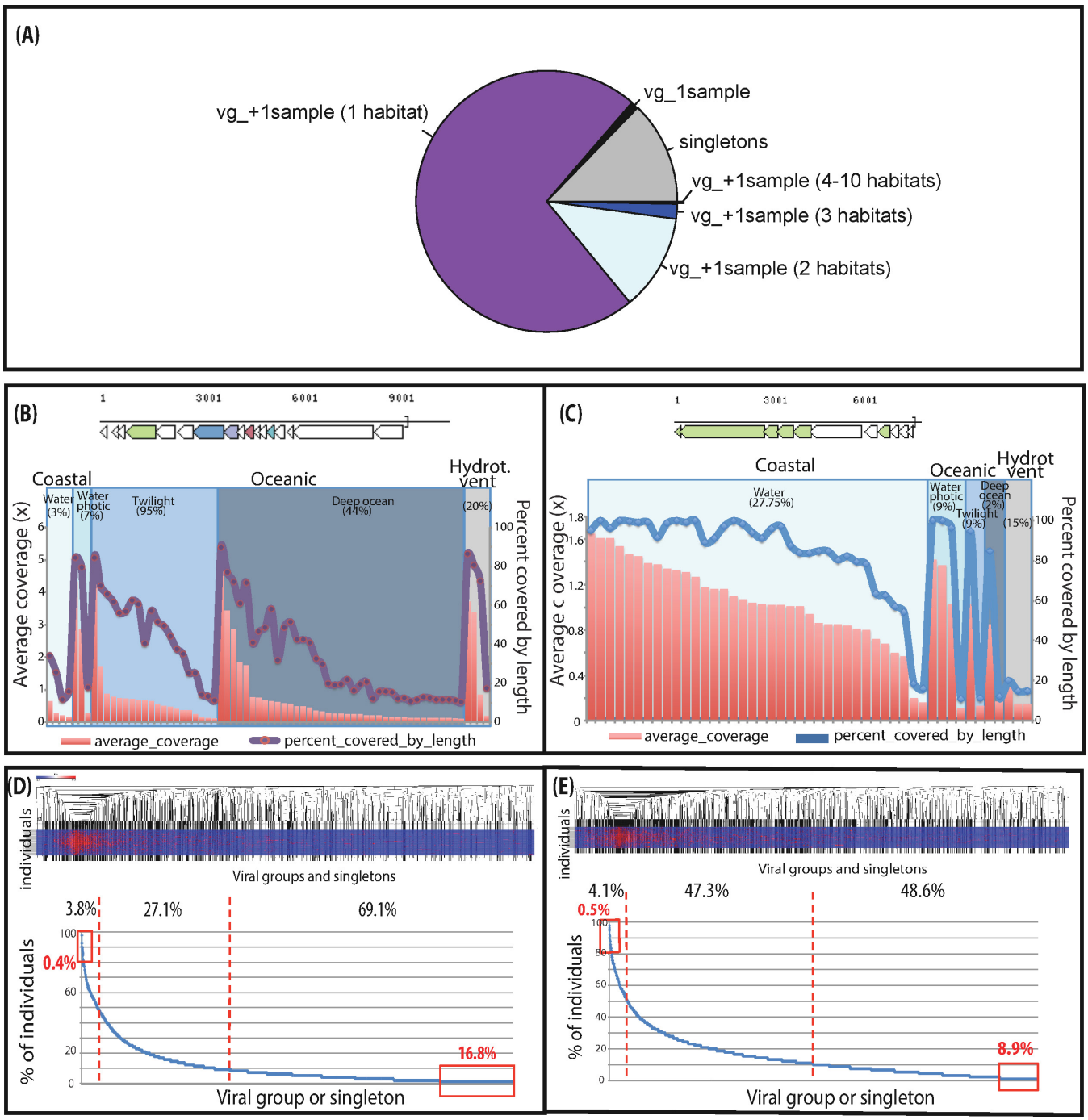
Extended Data Figure 5 | Viral group clustering method. a, Parameters used in the clustering of viral sequences. We used all 5,042 reference isolate viral genomes (iVGs) to group them using single-linkage hierarchical clustering (SLC) with different combinations of AAI and AF values to validate the clustering approach. The thresholds for AAI and AF were set at 90% and 50%, respectively, (boxed in purple) and were selected based on the accurate grouping of iVGs that was in agreement at the genus level, and the vast majority at the species level, according to the ICTV classification system (Supplementary Information). Further, these thresholds reduced the number of total connections (green line referred to secondary *y* axis) compared with lower AAI thresholds, without altering the total number of singletons and viral groups created (red and light green bars referred to primary *y* axis, respectively), as well as the average number of members per viral group (shown at the bottom of the figure). **b**, Size distribution of viral groups. Distribution of the 66,696 viral genomes clustered into 18,470 viral groups. Number of viral members (spanning from 2 to 365) per viral groups is shown. **c**–**e**, The cytoscape visualization of some viral groups. **c**, Major reference isolated viral groups created using SLC with AAI and AF values of 90% and 50%, respectively. Cytoscape force-directed (unweighted) layout option was used to visualize these groups. Black nodes represent isolated viral genomes whereas orange and green nodes represent metagenomic viral contigs clustered with isolates from host-associated and environmental samples, respectively. Group edges connect viral groups based on the above cutoffs. **d**, The four largest viral groups created from metagenomic viral contigs (containing 365, 201, 165, and 152 members, respectively). Specific habitat information of the samples as well as the viral group identifier is shown in the figure. **e**, Examples of viral groups (vg_2932 and vg_2864) containing proto-spacers (indicated by green circles) found in the CRISPR–Cas system of the indicated bacterial taxon. All the metagenomic viral contigs clustered in both viral groups were found in the same habitat subtype: human oral samples for vg_2932, and human faecal samples for vg_2864 (with a sole exception in the latter group that derived from an oral sample, indicated with a red arrow).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



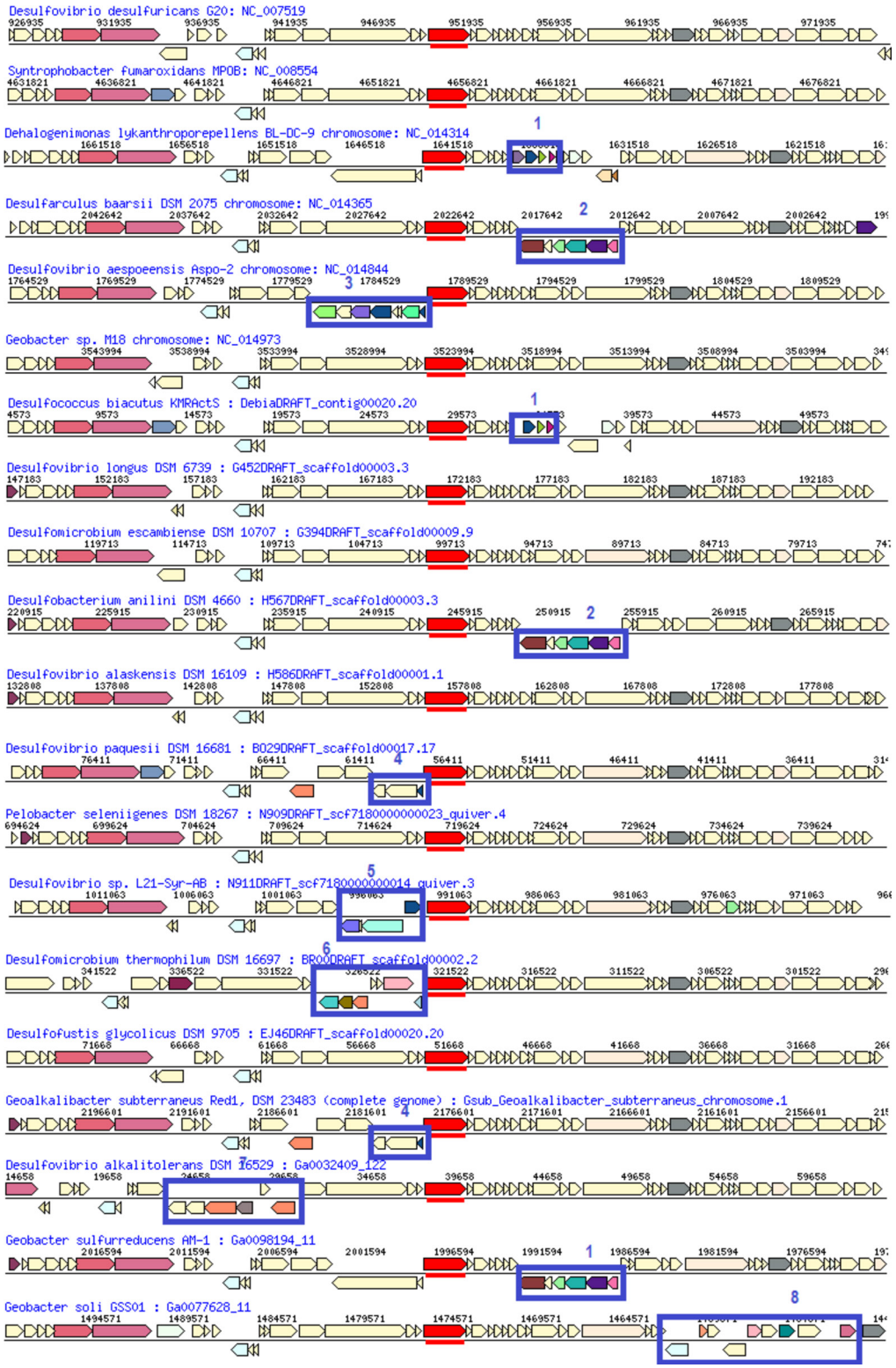
Extended Data Figure 6 | Verification of viruses identified with broad-host range. a, **b**, Alignments of all contigs found in the IMG database containing any of the 3 spacer matches present in a viral group potentially infecting 2 different phyla or any of the 7 spacer matches present in a viral group potentially infecting 3 different families are shown in **a** and **b**, respectively. Alignments were performed by mapping all the matches (48 for **a**, and 128 for **b**; named with an IMG scaffold identifier) to a viral representative using the 'map to reference' package of Geneious software (http://www.geneious.com). Black lines represent 100% sequence identity to the reference virus. The location of the 3 spacers (that derived from 2 different phyla) in **a** as well as the 7 spacers (that derived from 3 different families) in **b** is indicated with triangles with different colours. Spacer sequences, as well as the genomes that contain them in a CRISPR locus is boxed at the bottom.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



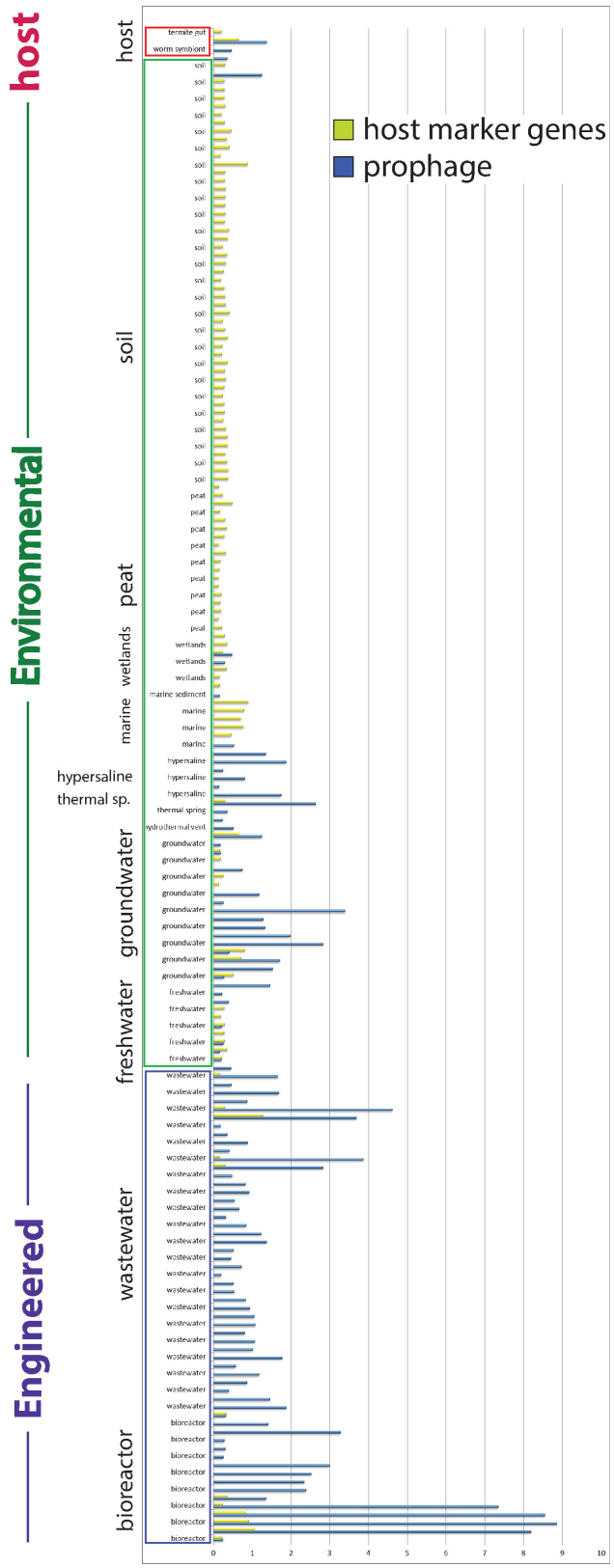
Extended Data Figure 7 | Habitat type specificity of all viral diversity and specific examples. a, Distribution of the presence of the total viral diversity of metagenomic viral contigs (viral groups and singletons) across distinct number of habitat types. A total of 85.9% of all viral diversity resided in a single habitat type (either as a singleton 19.7%, as a viral group found in a single sample 1.8%, or as a viral group found in 2 or more samples 64.4%), whereas only a small fraction (0.31% of all mVCs) were found in 4 or more different habitat types. **b**, **c**, Examples of viral groups found in diverse samples across different oceanic zones and provinces. Presence of a single viral group across distinct marine samples based on average coverage values (red bars; *y* axis on the left) and total percentage of the viral sequence length recovered per sample (purple line; *y* axis on the right). Samples were grouped by marine zones and indicate the percentage of the total samples per zone. **b**, Representative of viral group 2463 (IMG taxon id: 3300001450 and IMG scaffold id: JGI24006J15134_100002847) was found exclusively in marine biomes at depth and with reduced exposure to sunlight (across 95% of all twilight samples and in 44% of deep ocean samples). **c**, Representative of viral group 10643 (IMG taxon id: 3300000216 and IMG scaffold id: SI53jan11_150mDRAFT_c1002499) detected preferentially across coastal water samples (28% of all samples of this zone, preferentially in oxygen minimum zones), but also present in twilight, deep ocean, and hydrothermal vent samples. This viral group was identified as a SUP05-infecting phage. The genes of the viral contig representatives were coloured by the phylogenetic distribution of the best hit in the database (white, unknown; green, Proteobacteria; blue, Chlorophyta, red, unclassified virus; turquoise, Firmicutes; purple, Deinococcus). **d**, **e**, The distribution of viral sequences of distinct body sub-sites across different individuals. Hierarchical clustering (average linkage using Jaccard distance) was used for both axes (samples and individuals) across 'large intestine' (**d**) and 'oral' metagenomes (**e**), respectively (top chart in both panels). Presence or absence of viral groups or singletons per sample is colour-coded as red or blue, respectively. The line chart of both panels show the percentage of viral sharing for >50%, 50–10%, and <10% of the individuals (vertical lines) highlighting in red boxes the percentage of viral sharing for >80% as well as viral sequences only present in a single individual.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



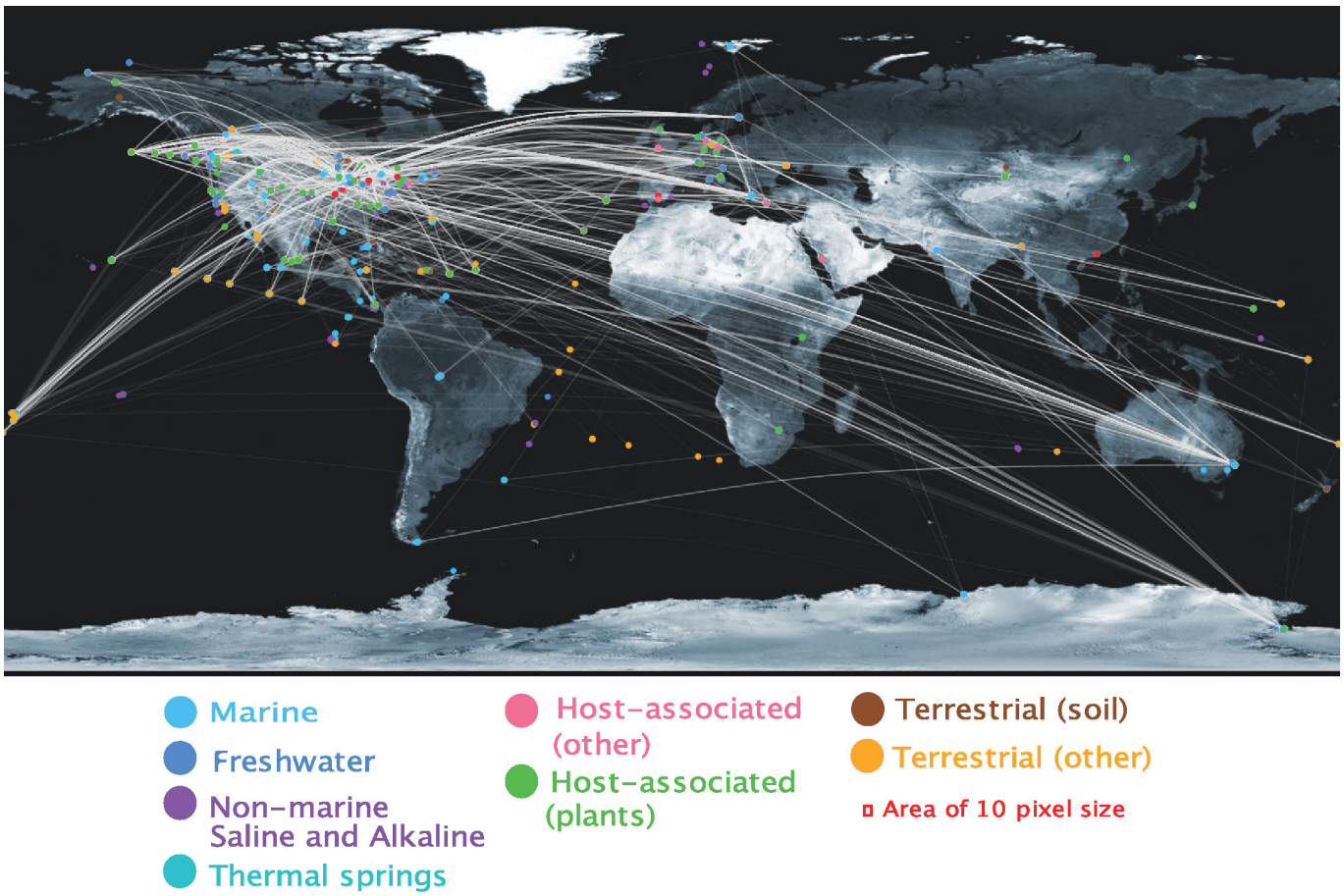
Extended Data Figure 8 | Alignment of broad-host specificity prophage in 20 isolate genomes in IMG using 'Gene Neighborhood' tool. The gene 'adenine-specific DNA methyltransferase' is used as an anchor for the alignment (in red). Genes are coloured according to COG cluster annotation, with light yellow representing genes without COG assignment. Blue boxes highlight likely cargo genes inserted in prophage genomes. These include: (1) alkyl hydroperoxide reductase system in *Dehalogenimonas lykanthroporepellens, Desulfococcus biacutus* and *Geobacter sulfurreducens*, (2) efflux ABC transporter in *Desulfoarculus baarsii* and *Desulfobacterium anilini*, (3) possible secondary metabolite biosynthesis genes in *Desulfovibrio aespoenensis*, (4) restriction system in *Desulfovibrio paquesii* and *Geoalkalibacter subterraneous*, (5) methionine synthase in *Desulfovibrio sp. L21-Syr-AB*, (6) molybdate ABC transporter in *Desulfomicrobium thermophilum*, (7) ABC transporter involved in multi-copper enzyme maturation in *Desulfovibrio alkalitolerans*; and (8) likely antibiotic resistance cassette in *Geobacter soli*. Details in Supplementary Table 24.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 9 | Distribution of hits to broad-host prophage and its potential hosts in metagenomic samples. The hits to prophage sequences and host marker genes (RNA polymerase subunits and ribosomal proteins) were identified by BLASTn with *e*-value 1.0×10^{50}; 90% nucleotide identity and cumulative alignment length of at least 10% of the length of the prophage or concatenated marker genes. Metagenome samples grouped by habitat are shown on the *y* axis; boxes correspond to broad environmental categories. Red box surrounds non-human host-associated samples (worm and termite symbionts), green box surrounds environmental samples (aquatic and terrestrial), blue box surrounds engineered samples (wastewater and bioreactors). Average coverage of the prophage and concatenated host marker genes is plotted on the *x* axis.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 10 | Global connectivity of viral diversity from different habitat types. Geographic location of metagenomic samples containing the same viral groups and singletons represented by a white connecting line across metagenomes from different habitats. Only samples sharing 2 or more viral groups or singletons that are more distant than 10 pixels (area shown as a red square in the figure) are connected. The colours of the samples (circles) indicate the habitat type according with the legend. A freely available equirectangular projection of the world map was used as a background image (http://visibleearth.nasa.gov/view.php?id=57752).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

ARTICLE

doi:10.1038/nature19083

Molecular basis of APC/C regulation by the spindle assembly checkpoint

Claudio Alfieri[1]*, Leifu Chang[1]*, Ziguo Zhang[1], Jing Yang[1], Sarah Maslen[1], Mark Skehel[1] & David Barford[1]

In the dividing eukaryotic cell, the spindle assembly checkpoint (SAC) ensures that each daughter cell inherits an identical set of chromosomes. The SAC coordinates the correct attachment of sister chromatid kinetochores to the mitotic spindle with activation of the anaphase-promoting complex (APC/C), the E3 ubiquitin ligase responsible for initiating chromosome separation. In response to unattached kinetochores, the SAC generates the mitotic checkpoint complex (MCC), which inhibits the APC/C and delays chromosome segregation. By cryo-electron microscopy, here we determine the near-atomic resolution structure of a human APC/C–MCC complex (APC/C^{MCC}). Degron-like sequences of the MCC subunit BubR1 block degron recognition sites on Cdc20, the APC/C coactivator subunit responsible for substrate interactions. BubR1 also obstructs binding of the initiating E2 enzyme UbcH10 to repress APC/C ubiquitination activity. Conformational variability of the complex enables UbcH10 association, and structural analysis shows how the Cdc20 subunit intrinsic to the MCC (Cdc20^{MCC}) is ubiquitinated, a process that results in APC/C reactivation when the SAC is silenced.

The fidelity of chromosome separation at each cell division cycle ensures the inheritance of the correct complement of genetic material in successive generations of cells. The APC/C (also known as the cyclosome) initiates sister chromatid separation by controlling the proteasomal degradation of securin and cyclin B[1,2]. Their degradation allows separase to remove sister chromatid cohesin. Crucial to the maintenance of chromosome integrity of dividing cells is the SAC[3,4]. The SAC responds to unattached kinetochores by generating the MCC, which functions to suppress APC/C-catalysed ubiquitination of securin and cyclin B.

Although the components of the SAC machinery are known[5,6], and some details of the molecular events that sense the absence of kinetochore attachment (and possibly intra-kinetochore tension) to signal MCC assembly have been characterized, important questions still remain[4]. Intrinsic to this process is the conversion of open (O-) to closed (C-) Mad2, catalysed by a C-Mad2–Mad1 complex at unattached kinetochores[7–9]. Soluble C-Mad2 engages the N terminus of Cdc20 (refs 10, 11), the mitotic activating subunit of the APC/C, which then binds the BubR1–Bub3 dimer to form the MCC[12]. Mad2 and BubR1 (also known as Mad3) interact cooperatively with Cdc20 (refs 9, 13–18), and synergistically inhibit the APC/C during mitosis[14,19].

In an important advance, it was proposed[2], and shown[20], that the tetrameric MCC inhibits the APC/C already in complex with the regulatory subunit Cdc20 (APC/C^{Cdc20}), which recognizes destruction box (D-box), KEN box and ABBA motif degrons of APC/C substrates and promotes the catalytically active conformation of the APC/C[21,22]. In addition to inhibiting the APC/C, the MCC contributes to APC/C reactivation after SAC silencing through proteasome-catalysed Cdc20 degradation[23,24]. SAC-mediated Cdc20 proteolysis is dependent on the APC/C, Mad2 and BubR1 (refs 17,23–27), suggesting that Cdc20 ubiquitination occurs in the context of the APC/C^{MCC}, an idea supported by findings showing that release from mitotic arrest, concomitant with Cdc20 destruction, requires the small APC/C subunit Apc15 (refs 28–30).

To obtain insights into reciprocal APC/C and MCC regulation, we reconstituted recombinant complexes of APC/C^{MCC} and APC/C^{MCC} plus the APC/C initiating E2 enzyme UbcH10 for structural and biochemical analysis. From a cryo-electron microscopy (cryo-EM) reconstruction of the APC/C^{MCC}, we identify conformational variability of the complex that explains its capacity to repress substrate ubiquitination, but also enables UbcH10 to catalyse intramolecular Cdc20^{MCC} ubiquitination.

Reconstitution and overall features of APC/C^{MCC}

We reconstituted recombinant APC/C^{MCC} using the insect cell/baculovirus expression system. Recombinant APC/C^{MCC} incorporates two distinct Cdc20 subunits, termed Cdc20$^{APC/C}$ and Cdc20^{MCC} for the APC/C^{Cdc20}-associated and MCC-associated subunits, respecitvely (Extended Data Fig. 1a, j and Extended Data Table 1), consistent with previous findings[20]. We determined negative-stain and cryo-EM reconstructions of the APC/C^{MCC} complex (Extended Data Table 2). The negative-stain EM reconstruction of the recombinant APC/C^{MCC} is essentially identical in structure to endogenous APC/C^{MCC} isolated from checkpoint-arrested HeLa cells determined at a similar resolution[31] (Extended Data Fig. 2b). This substantiates the model that the physiological form of APC/C^{MCC} includes two Cdc20 subunits[2,20]. Both reconstructions feature a large density element termed the MCC–Cdc20$^{APC/C}$ module (MCC interacting with the Cdc20$^{APC/C}$ subunit of APC/C^{Cdc20}) occupying the APC/C central cavity, extending from the 'front' side of the platform domain (Extended Data Fig. 2b).

To understand quantitatively how the MCC interacts with APC/C^{Cdc20}, we determined a cryo-EM reconstruction of APC/C^{MCC} at near-atomic resolution (Fig. 1a, b, Extended Data Figs 2c–e, 3 and Extended Data Table 2). Extensive 3D classification of APC/C^{MCC} revealed conformational variability of the MCC–Cdc20$^{APC/C}$ module. This module adopts a stable, rigid conformation in 21% of APC/C^{MCC} particles (defined as APC/C$^{MCC\text{-}closed}$) (Extended Data Fig. 4b, class 1). A local resolution map of APC/C$^{MCC\text{-}closed}$ shows that the central rigid regions are at 3.9–4.1 Å resolution, with the flexible outer regions at lower resolution (Extended Data Fig. 3c, d). We built an atomic model of APC/C$^{MCC\text{-}closed}$ (Fig. 1 and Supplementary Video 1) guided by the cryo-EM structure of human APC/C$^{Cdh1.Emi1}$ (ref. 32) and the crystal structure of fission yeast MCC[18].

In the APC/C$^{MCC\text{-}closed}$ reconstruction, the secondary structure elements of the MCC–Cdc20$^{APC/C}$ module are clearly visible

[1]MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, UK.
*These authors contributed equally to this work.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

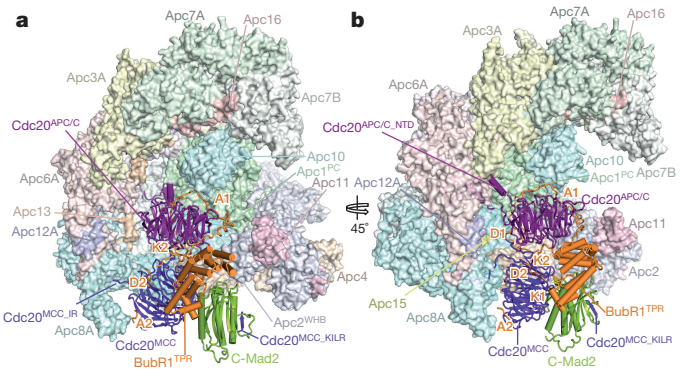


Figure 1 | Overall structure of the APC/C^{MCC} complex. a, **b**, Two views of APC/C^{MCC}. The MCC–Cdc20$^{APC/C}$ module is shown as a cartoon and the APC/C in a surface representation. BubR1 forms extensive contacts with Cdc20$^{APC/C}$ and Apc2. BubR1 inhibitory degrons visible in these views are highlighted.

(Extended Data Fig. 2e), revealing that Cdc20$^{APC/C}$, Cdc20^{MCC}, Mad2 and the N-terminal and middle regions of BubR1, including its tetratricopeptide repeat (TPR) domain (BubR1^{TPR}) and its inhibitory degrons, are well defined (Figs 1a, b and 2). However, despite their presence in reconstituted APC/C^{MCC} (Extended Data Figs 1a and 2a), the C-terminal pseudo-kinase domain of BubR1 (Fig. 2a) and Bub3, were not visible in EM density, indicative of conformational variability. In agreement with this, both the structure and activity of APC/C^{MCC} with Bub3 and the BubR1 C terminus deleted (APC/C^{miniMCC}) are indistinguishable from APC/C^{MCC} (Extended Data Figs 1b, h, i and 2b). These data are consistent with the requirement of the N-terminal ~363 residues of BubR1 (Fig. 2a) to sustain a SAC[33–36].

General architecture of APC/C^{MCC}

In APC/C^{MCC}, the MCC core elements comprising Cdc20^{MCC}, BubR1^{TPR} and Mad2 resemble their counterparts in *Schizosaccharomyces pombe* MCC[18] (Fig. 1 and Supplementary Video 1). Mad2 adopts the closed conformation (C-Mad2) with its C-terminal segment ('safety belt') locking the tubular density of the Cdc20^{MCC} KILR motif[10,11] (Figs 1, 2c and Extended Data Fig. 5c). The MCC docks into the APC/C^{Cdc20} cavity directly below Cdc20$^{APC/C}$, and interacts with Cdc20$^{APC/C}$ such that the two Cdc20 WD40 domains of APC/C^{MCC} are arranged in an almost perpendicular fashion (Figs 1 and 2c). Cdc20^{MCC} and BubR1 mediate interactions between the MCC and APC/C^{Cdc20}, with BubR1 dominating these interactions through its contacts to Cdc20$^{APC/C}$ and Apc2. C-Mad2 forms no direct contacts with APC/C^{Cdc20} (Fig. 1). However, by stabilizing the association of BubR1 with Cdc20^{MCC}, C-Mad2 indirectly augments BubR1 interactions with APC/C^{Cdc20} (refs 14, 19, 35). APC/C$^{MCC\text{-}closed}$ is similar in structure to APC/C and coactivator complexes[22,32], with major conformational differences confined to Cdc20$^{APC/C}$ and the platform subunits Apc4, Apc5 and Apc15 (Extended Data Fig. 5d, e).

Contacts between the two Cdc20 subunits of APC/C$^{MCC\text{-}closed}$ are mainly mediated by BubR1 that intertwines between them (Fig. 2c). In response to MCC binding, the WD40 domain of Cdc20$^{APC/C}$ is tilted by ~40° and rotated 90° about its central axis[22] (Extended Data Fig. 5d). This disrupts the D-box-binding site formed by the interface of the D-box co-receptors of Cdc20$^{APC/C\text{-}WD40}$ and Apc10 (refs 21, 32), and pulls the C-terminal isoleucine-arginine motif (IR tail) of APC/C^{Cdc20} away from the IR-tail-binding site of Apc3A. A disengagement of the Cdc20$^{APC/C}$ IR tail from Apc3A is consistent with weak EM density at the IR-tail-binding site (Extended Data Fig. 5a), and the finding that Cdc20 does not require Apc3 to bind the APC/C when the SAC is active[37]. In contrast to the IR tail, the N-terminal domain of Cdc20$^{APC/C}$ maintains the same interactions with Apc8B and Apc1^{PC} as seen in APC/C^{Cdc20} (ref. 22) (Fig. 3a and Extended Data Fig. 5b).

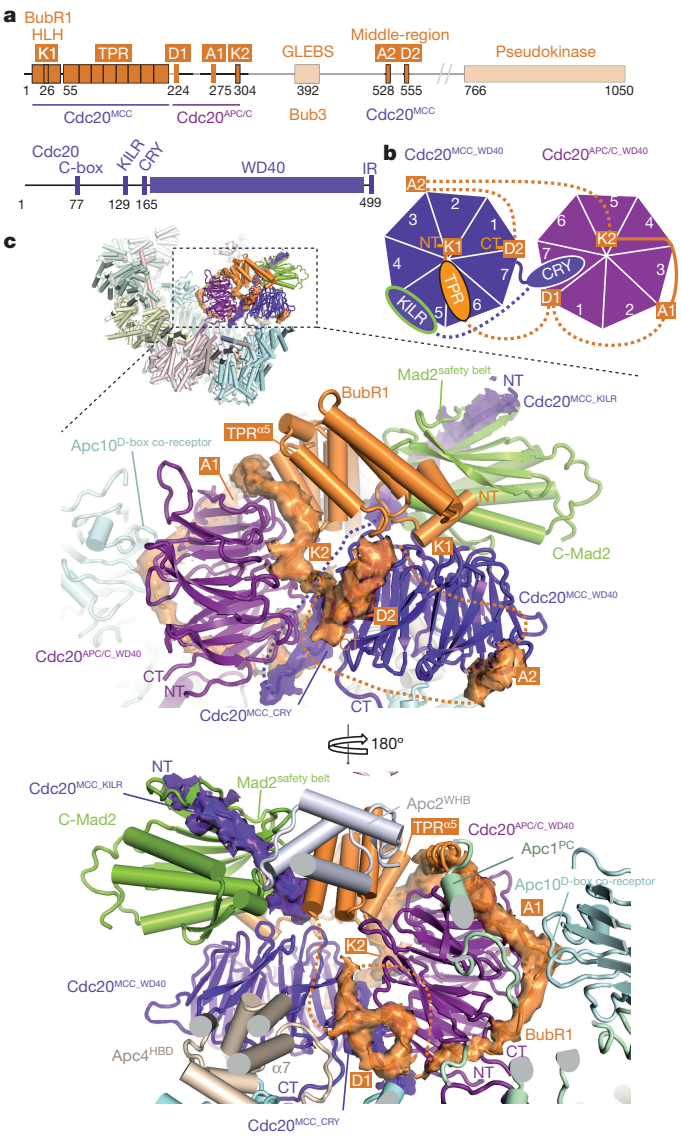


Figure 2 | Interactions of BubR1 with Cdc20$^{APC/C}$ and Cdc20^{MCC}. **a**, Schematics of BubR1 (top) and Cdc20 (bottom). **b**, Schematic representation of the top views of the Cdc20^{APC} and Cdc20^{MCC} WD40 domains. WD40 domain blades are numbered and the positions of BubR1 inhibitory degrons (orange) are indicated. The CRY degron[49] mediates Cdc20^{MCC} interactions with Cdc20$^{APC/C}$ (Extended Data Fig. 5f). **c**, Two views showing details of the MCC–Cdc20$^{APC/C}$ module. Cryo-EM density of the BubR1 inhibitory degrons, Cdc20^{MCC} CRY box and KILR motif is shown. Interactions of the BubR1 A1 motif with the Apc10$^{D\text{-box co-receptor}}$ and Apc1; BubR1^{TPR} with Apc2^{WHB}; and Cdc20^{MCC} with Apc4^{HBD} are indicated (bottom). Inset, overall view of APC/C^{MCC}.

BubR1 interacts with both Cdc20 subunits of APC/C^{MCC}

Similar to yeast MCC[18], in APC/C^{MCC}, the N-terminal KEN-1 box (K1) (Fig. 2a and Supplementary Video 1) engages the KEN-box recognition site of the Cdc20^{MCC} (Fig. 2b, c). Immediately after BubR1^{TPR} and preceding KEN-2 (K2) is the N-terminal D-box motif (D1)[20] (Fig. 2a). We assigned D1 to the loop-like density at the D-box recognition site of Cdc20$^{APC/C}$, although owing to the re-orientation of Cdc20$^{APC/C}$ in APC/C^{MCC}, its position is diametrically opposed to its location in active APC/C–coactivator complexes[21,22,32] (Extended Data Fig. 5d). D1 and its flanking residues mediate the interface of Cdc20$^{APC/C}$ and Cdc20^{MCC} (Figs 1 and 2c, bottom). This explains how mutating D1 disrupts MCC binding to a second Cdc20 subunit without affecting MCC integrity[20]. By engaging the Cdc20$^{APC/C}$ D-box co-receptor in APC/C^{MCC}, D1 would obstruct D-box-dependent

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

substrate recognition, consistent with its requirement for a checkpoint arrest[20].

EM density immediately C-terminal to D1 is disordered (Fig. 2c, bottom); however, uninterrupted tubular density situated nearby and assigned to BubR1 is clearly defined wrapping around the opposite side of the Cdc20$^{APC/C}$ WD40 domain, connecting its bottom and top surfaces and contacting both the Cdc20$^{APC/C}$ KEN and ABBA motif-binding sites (labelled K2 and A1 in Fig. 2c). This BubR1 EM density feature bears a notable resemblance to the structure of Acm1 (APC/C-Cdh1-modulator 1), an inhibitor of yeast Cdh1 that uses ABBA, D-box and KEN motifs to block the degron-recognition sites of Cdh1 (ref. 38) (Extended Data Fig. 6a). The BubR1 EM density contacting the top surface of Cdc20$^{APC/C}$ corresponds to the KEN-2 box (Fig. 2c, top). Guided by the Cdh1–Acm1 crystal structure[38], we modelled the KEN-2 box and the preceding ABBA motif (Fig. 2a and Extended Data Fig. 6a, b) to the KEN box and ABBA motif recognition sites of Cdc20$^{APC/C}$ (Fig. 2c).

The KEN-2 motif is invariant in BubR1 orthologues (Extended Data Fig. 6b), and although not required for MCC assembly[16,18,20,26,36,39], it is essential for a spindle checkpoint arrest[16,26,36,39]. Thus, similar to D1, a role for KEN-2 in mediating MCC interactions with APC/C^{Cdc20} would explain its requirement in the SAC by stabilizing MCC interactions with Cdc20$^{APC/C}$ (ref. 20) and inhibiting degron recognition by APC/C^{Cdc20}.

A middle segment of metazoan BubR1 includes an ABBA motif (A2) and a D-box (D2) (Fig. 2a) that mediate Cdc20–BubR1 interactions in a Mad2-independent manner[15,19,40–43]. These motifs contribute weakly to sustaining the SAC[15,41,42], and have a role in recruiting Cdc20 to unattached kinetochores[40–42]. The tubular EM densities located at the ABBA motif and D-box recognition sites of Cdc20^{MCC} were assigned to these motifs (Fig. 2c, top).

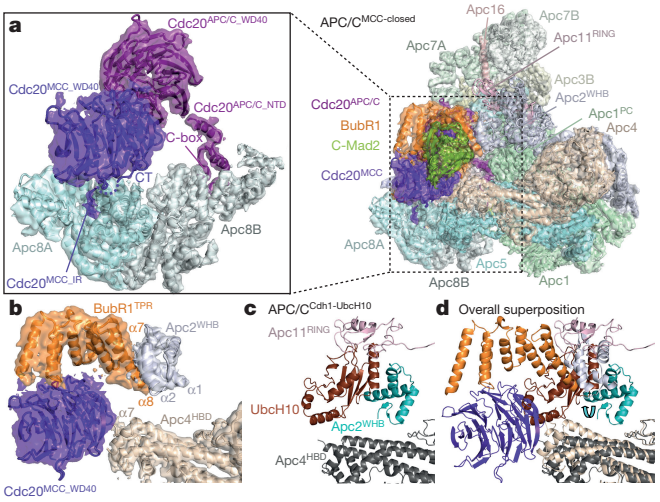


Figure 3 | Interactions of the MCC–Cdc20$^{APC/C}$ module with the APC/C, and APC/C catalytic inhibition by the MCC. a, Right, an overview of the APC/C^{MCC} model with the corresponding cryo-EM density. Left, segmented cryo-EM density of the Apc8A–Apc8B dimer and its two associated Cdc20 molecules. Cdc20$^{APC/C}$ interacts with Apc8B via its N-terminal domain (NTD). Cdc20^{MCC} interacts with Apc8A through its C-terminal Ile-Arg (IR) tail. **b**–**d**, Comparison of the binding mode of BubR1 and Cdc20^{MCC_WD40} in APC/C^{MCC} with the binding mode of UbcH10 in APC/C$^{Cdh1\text{-}UbcH10\text{-}Ub}$ (ref. 32). **b**, Segmented cryo-EM density of Cdc20^{MCC}, BubR1^{TPR}, Apc4^{HBD} and Apc2^{WHB}. **c**, APC/C$^{Cdh1\text{-}UbcH10\text{-}Ub}$. **d**, Both structures were superposed. BubR1^{TPR} and Cdc20^{MCC_WD40} compete for the same binding surfaces on Apc2^{WHB} and Apc4^{HBD} that form the UbcH10-binding site in APC/C$^{Cdh1\text{-}UbcH10\text{-}Ub}$ (Extended Data Fig. 5g). The Apc2^{WHB} sub-domain of Apc2 is shifted in the APC/C^{MCC} complex relative to APC/C$^{Cdh1\text{-}UbcH10\text{-}Ub}$ and would clash with the UbcH10-binding site on Apc11^{RING}.

Cdc20^{MCC} contacts to the APC/C

Although Cdc20^{MCC} interactions with the APC/C are mainly mediated through BubR1 and APC/C^{Cdc20}, two additional contacts are notable. First, the C-terminal IR tail of Cdc20^{MCC} binds to a site on Apc8A that is structurally equivalent to the Cdc20$^{APC/C}$ C-box binding site on Apc8B[32] (Fig. 3a and Supplementary Video 1). This interaction can be rationalized from the similarities in how the Apc3 IR tail and Apc8 C-box binding sites interact with their cognate IR tail and C-box motifs, respectively[32]. Second, the lower surface of the Cdc20^{MCC} WD40 domain interacts with the conserved acidic region of the Apc4 helix-bundle domain (Apc4^{HBD}), close to the UbcH10-binding site[32,44] (Figs 2c and 3a, b).

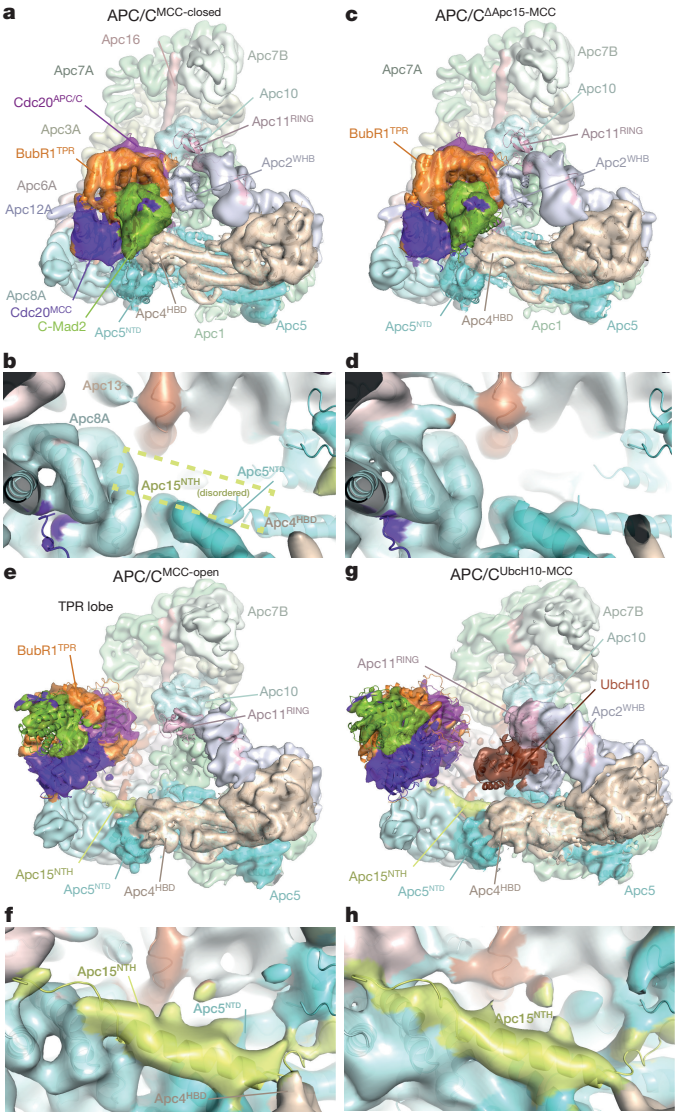


Figure 4 | Cryo-EM structures of APC/C$^{MCC\text{-}open}$, APC/C$^{\Delta Apc15\text{-}MCC}$ and APC/C$^{UbcH10\text{-}MCC}$, and comparison with APC/C$^{MCC\text{-}closed}$. a, Overall view of the cryo-EM density of APC/C$^{MCC\text{-}closed}$ and fitted coordinates for the MCC–Cdc20$^{APC/C}$ module, Apc2^{WHB}, Apc11, Apc4 and Apc5. The APC/C^{MCC} subunits are coloured as in Figs 1 and 3. **b**, Details of the Apc15^{NTH}-binding site on Apc8A and Apc5. Apc8A and Apc5 are shown. The position of the disordered Apc15^{NTH} is indicated by a box. **c**, **d**, APC/C$^{\Delta Apc15\text{-}MCC}$ complex (Apc15 deleted). **e**, **f**, APC/C$^{MCC\text{-}open}$ complex. **g**, **h**, APC/C$^{UbcH10\text{-}MCC}$ complex. In APC/C$^{MCC\text{-}closed}$ (**a**, **b**), BubR1^{TPR} interacts with Apc2^{WHB}, and Apc15^{NTH} is disordered. In APC/C$^{\Delta Apc15\text{-}MCC}$ (**c**, **d**), the MCC–Cdc20$^{APC/C}$ adopts the closed conformation, blocking the catalytic module. Conversely in APC/C$^{MCC\text{-}open}$ (**e**, **f**) and in APC/C$^{UbcH10\text{-}MCC}$ (**g**, **h**), BubR1^{TPR} no longer interacts with Apc2^{WHB}, and Apc15^{NTH} is ordered. **g**, In APC/C$^{UbcH10\text{-}MCC}$, Apc2^{WHB} and Apc11^{RING} interact with UbcH10. All cryo-EM reconstructions were filtered to 8.5 Å.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

The MCC suppresses APC/C E3 ligase activity

UbcH10 interacts with the APC/C through the RING domain of Apc11 and the winged-helix B (WHB) domain of Apc2 (Apc2^{WHB})[32,44]. In APC/C$^{MCC\text{-}closed}$, BubR1^{TPR} occludes both E2-binding sites (Figs 1 and 3). BubR1^{TPR} interacts directly with the UbcH10 interface of Apc2^{WHB} that repositions to contact BubR1^{TPR} (refs 32, 44) (Fig. 3b–d and Extended Data Figs 5g and 6c). Thus, in APC/C$^{MCC\text{-}closed}$, the steric occlusion of UbcH10 binding would inhibit ubiquitin chain synthesis.

Apc15-dependent conformational flexibility of the MCC

The APC/C$^{MCC\text{-}closed}$ structure explains how the MCC inhibits APC/C^{Cdc20} by obstructing substrate and UbcH10 interactions. However, Cdc20 auto-ubiquitination is the trigger for the spontaneous reactivation of the APC/C[23–29,45], dependent on the small APC/C subunit Apc15 (refs 28, 29). In previously reported APC/C structures[22,32], Apc15 adopts an extended conformation, anchored to Apc5 by its N terminus, and bridging Apc5 and Apc8A through its adjacent N-terminal helix (Apc15^{NTH}). Notably, in APC/C$^{MCC\text{-}closed}$, Apc15^{NTH} is structurally disordered (Fig. 4a, b). This results from Cdc20^{MCC} shifting the tip of Apc4^{HBD} that concomitantly repositions the adjacent N-terminal domain of Apc5 (Apc5^{NTD}), disrupting its interaction with Apc15^{NTH} (Fig. 4a, b, Extended Data Fig. 5d, e and Supplementary Video 2). The disorder of Apc15^{NTH} and repositioning of Apc4^{HBD} and Apc5^{NTD} in APC/C$^{MCC\text{-}closed}$ contrasts with their conformations in another structural state representing only 2% of APC/C^{MCC} particles (termed APC/C$^{MCC\text{-}open}$) (Methods, Extended Data Fig. 4a, b and Extended Data Table 2). In APC/C$^{MCC\text{-}open}$, EM density assigned to the MCC–Cdc20$^{APC/C}$ module is weaker relative to APC/C$^{MCC\text{-}closed}$. However, EM density for Apc15^{NTH} is clearly defined (Fig. 4e, f). Moreover, in APC/C$^{MCC\text{-}open}$, MCC–Cdc20$^{APC/C}$ is shifted towards the TPR lobe of APC/C, disrupting contacts between the MCC and the catalytic module, Apc4^{HBD} and Apc10. The 'open' position of MCC–Cdc20$^{APC/C}$ is stabilized by Apc15^{NTH} interacting with Apc5^{NTD} that pushes onto the Cdc20^{MCC}-binding site of Apc4^{HBD} (Fig. 4f and Extended Data Fig. 5e). Thus, the stability of Apc15^{NTH} influences the transition between open and closed conformations of the MCC–Cdc20$^{APC/C}$ module.

To investigate further how Apc15 modulates the position of MCC–Cdc20$^{APC/C}$, we determined the structure of APC/C^{MCC} with Apc15 deleted (APC/C$^{\Delta Apc15}$) (Extended Data Table 2, Fig. 4c, d and Extended Data Fig. 3). APC/C$^{\Delta Apc15}$ assembled similarly to wild-type APC/C and was catalytically active towards securin (Extended Data Fig. 1k, l). However, in contrast to APC/C^{MCC}, the reconstituted APC/C$^{\Delta Apc15\text{-}MCC}$ was defective for Cdc20 auto-ubiquitination, consistent with previous reports[28,29] (Fig. 5c and Extended Data Fig. 1h). 3D classification of the APC/C$^{\Delta Apc15\text{-}MCC}$ cryo-EM data set showed that the

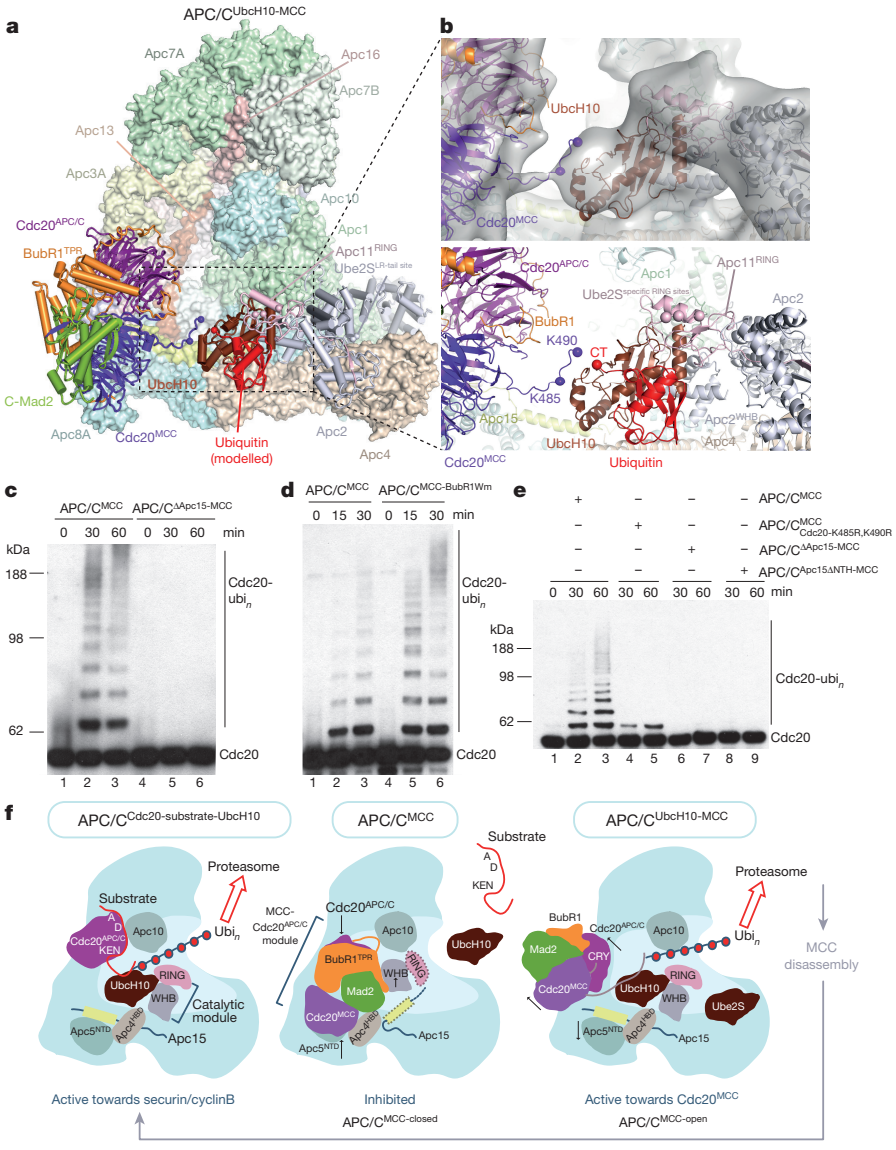


Figure 5 | Mechanism of Cdc20 auto-ubiquitination by APC/C$^{UbcH10\text{-}MCC}$. **a**, **b**, Model of a Cdc20^{MCC} ubiquitination complex based on the APC/C$^{UbcH10\text{-}MCC}$ cryo-EM reconstruction. The UbcH10–ubiquitin conjugate is modelled in the closed conformation[50]. **b**, Top, cryo-EM density and model of APC/C$^{UbcH10\text{-}MCC}$. The EM map is filtered to 12 Å and displayed at slightly lower threshold than in Fig. 4g, see Extended Data Fig. 3e, for comparisons. Clear EM density connects Cdc20^{MCC} with UbcH10. Bottom, the Cdc20^{MCC} pre-ubiquitination model. Cdc20^{MCC} residues Lys485 and Lys490, ubiquitinated in logarithmic and checkpoint-arrested cells, respectively[30], are in close proximity to the UbcH10 catalytic site (red sphere). **c**, Apc15 is required for Cdc20 ubiquitination by recombinant APC/C^{MCC}. **d**, BubR1^{Wm} mutations at the Apc2^{WHB} interface (R169A, F175A,V200A and L205) (Extended Data Fig. 5g) stimulate Cdc20 ubiquitination. **e**, Cdc20 residues Lys485 and Lys490 are ubiquitinated by recombinant APC/C^{MCC} (compare lanes 2, 3 and 4, 5). Apc15^{NTH} is required for Cdc20 ubiquitination by recombinant APC/C^{MCC} (lanes 8, 9). **f**, Cartoon illustrating reciprocal regulation of APC/C and MCC by APC/C^{MCC}. In APC/C$^{MCC\text{-}closed}$, MCC inhibits substrate (for example, securin and cyclin B) and UbcH10 recognition. Cyclin A and Nek2A can bypass the SAC. In APC/C$^{MCC\text{-}open}$, the UbcH10-binding site is exposed. Ube2S elongates Ub-conjugates initiated by UbcH10. Experiments in **c**–**e** were replicated three times. See Supplementary Fig. 1 for gel source data.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

MCC–Cdc20$^{APC/C}$ module adopts only the closed configuration (Fig. 4c, d, Extended Data Fig. 7b and Supplementary Video 2), with no evidence of APC/C$^{MCC\text{-}open}$ (Extended Data Fig. 7a, classes 1 to 6). Thus, the absence of Apc15 promotes APC/C$^{MCC\text{-}closed}$, explaining why Apc15 is necessary for Cdc20 ubiquitination. In support of our structural data that Apc15^{NTH} is required to stabilize APC/C$^{MCC\text{-}open}$, similar to APC/C$^{\Delta Apc15\text{-}MCC}$, APC/C^{MCC} reconstituted with Apc15 lacking the Apc15^{NTH} (Apc15$^{\Delta NTH}$) is defective in Cdc20 auto-ubiquitination (Fig. 5e, lanes 8 and 9, and Extended Data Fig. 1h).

Since interactions between Apc2^{WHB} and BubR1 stabilize APC/C$^{MCC\text{-}closed}$ (Figs 1 and 3b), disrupting this interface should favour APC/C$^{MCC\text{-}open}$. Consistent with this idea, negative-stain EM reconstructions of an APC/C^{MCC} mutant with Apc2^{WHB} deleted showed APC/C^{MCC} adopting the open conformation, with no APC/C$^{MCC\text{-}closed}$ (Extended Data Figs 1c and 2b). Importantly, in the complementary Apc2^{WHB}-binding surface mutant of BubR1 (Extended Data Fig. 5g), Cdc20^{MCC} auto-ubiquitination is stimulated (BubR1^{Wm}; Fig. 5d and Extended Data Fig. 1h).

Mechanism of Cdc20^{MCC} auto-ubiquitination

Cdc20^{MCC} auto-ubiquitination catalysed by APC/C^{MCC}–UbcH10 is an intra-molecular process reliant on the APC/C$^{MCC\text{-}open}$ conformation. To explore this possibility further, we determined the cryo-EM structure of APC/C^{MCC} in complex with UbcH10 (APC/C$^{UbcH10\text{-}MCC}$) (Extended Data Table 2). A 3D reconstruction using 7% of total particles showed clear EM density for both UbcH10 and the MCC–Cdc20$^{APC/C}$ module (Extended Data Fig. 7c). The resultant EM map at 8.9 Å resolution allowed rigid body docking of the APC/C, UbcH10 and MCC–Cdc20$^{APC/C}$ coordinates. APC/C$^{UbcH10\text{-}MCC}$ resembles APC/C$^{MCC\text{-}open}$ (Figs 4e–h, 5a, Extended Data Fig. 3 and Supplementary Video 2). Apc15^{NTH} is ordered and MCC–Cdc20$^{APC/C}$ is rotated towards the TPR lobe leaving the catalytic module accessible to bind UbcH10 (Fig. 5a).

Notably, in APC/C$^{UbcH10\text{-}MCC}$, UbcH10 induces an additional small rotation of MCC–Cdc20$^{APC/C}$. The C terminus of Cdc20^{MCC} is visualized as an extended tubular density feature linking the Cdc20^{MCC} WD40 domain with the catalytic site of UbcH10 (Fig. 5b, Extended Data Fig. 3e and Supplementary Video 2). Modelling the C terminus of Cdc20^{MCC} into this density shows that Lys485 and Lys490 are accessible to the UbcH10 catalytic site (Fig. 5b). Both residues are ubiquitinated *in vitro* (data not shown), and their replacement by Arg virtually eliminated Cdc20 auto-ubiquitination (Fig. 5e), indicating that these two residues are the major sites of Cdc20^{MCC} ubiquitination in the context of APC/C$^{MCC\text{-}UbcH10}$. Our model, that the MCC competitively restricts access of UbcH10 to its binding site on the APC/C, is consistent with the reduced binding of UbcH10 to endogenous APC/C^{MCC} relative to APC/C^{Cdc20} (ref. 31), and our finding that compared with securin ubiquitination, Cdc20 auto-ubiquitination required a tenfold higher concentration of UbcH10 (Extended Data Fig. 1m), is in agreement with previous findings[29]. Although our data show that MCC suppresses APC/C–UbcH10 interactions, the effect of the MCC on modulating the affinity of the APC/C for the elongating E2 Ube2S, which contributes to checkpoint silencing[46–48], is less clear. The binding site for the Ube2S C-terminal LR tail at the Apc2–Apc4 interface[32], necessary for its association with the APC/C, is unaffected by the MCC (Fig. 5a), thus in principle allowing Ube2S to assemble ubiquitin chains onto the C terminus of Cdc20^{MCC}. However, the capacity of Cdc20 to stabilize APC/C–Ube2S complexes[48] may be affected by the interactions of BubR1 with the APC/C^{MCC}.

In conclusion, our structural analysis of APC/C^{MCC} and APC/C$^{UbcH10\text{-}MCC}$ provides a molecular explanation for the spontaneous APC/C reactivation resulting from SAC inactivation involving ubiquitin-mediated Cdc20^{MCC} proteolysis. APC/C reactivation in response to the cessation of SAC signalling is facilitated by the reciprocal control of APC/C and MCC activities mediated by the conformational flexibility of APC/C^{MCC}, influenced by Apc15. Modulation of this conformational change could allow for the regulation of Cdc20^{MCC} auto-ubiquitination (Fig. 5f). A candidate for such a role is p31^{comet}, a Mad2-binding protein proposed to stimulate Cdc20 ubiquitination in checkpoint-arrested cells[23,45].

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 March; accepted 6 July 2016.
Published online 10 August 2016.

1. Meyer, H. J. & Rape, M. Processive ubiquitin chain formation by the anaphase-promoting complex. *Semin. Cell Dev. Biol.* **22,** 544–550 (2011).
2. Primorac, I. & Musacchio, A. Panta rhei: the APC/C at steady state. *J. Cell Biol.* **201,** 177–189 (2013).
3. Lara-Gonzalez, P., Westhorpe, F. G. & Taylor, S. S. The spindle assembly checkpoint. *Curr. Biol.* **22,** R966–R980 (2012).
4. Musacchio, A. The molecular biology of spindle assembly checkpoint signaling dynamics. *Curr. Biol.* **25,** R1002–R1018 (2015).
5. Hoyt, M. A., Totis, L. & Roberts, B. T. *S. cerevisiae* genes required for cell cycle arrest in response to loss of microtubule function. *Cell* **66,** 507–517 (1991).
6. Li, R. & Murray, A. W. Feedback control of mitosis in budding yeast. *Cell* **66,** 519–531 (1991).
7. De Antoni, A. *et al.* The Mad1/Mad2 complex as a template for Mad2 activation in the spindle assembly checkpoint. *Curr. Biol.* **15,** 214–225 (2005).
8. Mapelli, M., Massimiliano, L., Santaguida, S. & Musacchio, A. The Mad2 conformational dimer: structure and implications for the spindle assembly checkpoint. *Cell* **131,** 730–743 (2007).
9. Kulukian, A., Han, J. S. & Cleveland, D. W. Unattached kinetochores catalyze production of an anaphase inhibitor that requires a Mad2 template to prime Cdc20 for BubR1 binding. *Dev. Cell* **16,** 105–117 (2009).
10. Luo, X., Tang, Z., Rizo, J. & Yu, H. The Mad2 spindle checkpoint protein undergoes similar major conformational changes upon binding to either Mad1 or Cdc20. *Mol. Cell* **9,** 59–71 (2002).
11. Sironi, L. *et al.* Crystal structure of the tetrameric Mad1-Mad2 core complex: implications of a 'safety belt' binding mechanism for the spindle checkpoint. *EMBO J.* **21,** 2496–2506 (2002).
12. Sudakin, V., Chan, G. K. & Yen, T. J. Checkpoint inhibition of the APC/C in HeLa cells is mediated by a complex of BUBR1, BUB3, CDC20, and MAD2. *J. Cell Biol.* **154,** 925–936 (2001).
13. Hardwick, K. G., Johnston, R. C., Smith, D. L. & Murray, A. W. MAD3 encodes a novel component of the spindle checkpoint which interacts with Bub3p, Cdc20p, and Mad2p. *J. Cell Biol.* **148,** 871–882 (2000).
14. Fang, G. Checkpoint protein BubR1 acts synergistically with Mad2 to inhibit anaphase-promoting complex. *Mol. Biol. Cell* **13,** 755–766 (2002).
15. Davenport, J., Harris, L. D. & Goorha, R. Spindle checkpoint function requires Mad2-dependent Cdc20 binding to the Mad3 homology domain of BubR1. *Exp. Cell Res.* **312,** 1831–1842 (2006).
16. Burton, J. L. & Solomon, M. J. Mad3p, a pseudosubstrate inhibitor of APCCdc20 in the spindle assembly checkpoint. *Genes Dev.* **21,** 655–667 (2007).
17. Nilsson, J., Yekezare, M., Minshull, J. & Pines, J. The APC/C maintains the spindle assembly checkpoint by targeting Cdc20 for destruction. *Nat. Cell Biol.* **10,** 1411–1420 (2008).
18. Chao, W. C., Kulkarni, K., Zhang, Z., Kong, E. H. & Barford, D. Structure of the mitotic checkpoint complex. *Nature* **484,** 208–213 (2012).
19. Tang, Z., Bharadwaj, R., Li, B. & Yu, H. Mad2-Independent inhibition of APCCdc20 by the mitotic checkpoint protein BubR1. *Dev. Cell* **1,** 227–237 (2001).
20. Izawa, D. & Pines, J. The mitotic checkpoint complex binds a second CDC20 to inhibit active APC/C. *Nature* **517,** 631–634 (2015).
21. Chang, L., Zhang, Z., Yang, J., McLaughlin, S. H. & Barford, D. Molecular architecture and mechanism of the anaphase-promoting complex. *Nature* **513,** 388–393 (2014).
22. Zhang, S. *et al.* Molecular mechanism of APC/C activation by mitotic phosphorylation. *Nature* **533,** 260–264 (2016).
23. Reddy, S. K., Rape, M., Margansky, W. A. & Kirschner, M. W. Ubiquitination by the anaphase-promoting complex drives spindle checkpoint inactivation. *Nature* **446,** 921–925 (2007).
24. Stegmeier, F. *et al.* Anaphase initiation is regulated by antagonistic ubiquitination and deubiquitination activities. *Nature* **446,** 876–881 (2007).
25. Pan, J. & Chen, R. H. Spindle checkpoint regulates Cdc20p stability in *Saccharomyces cerevisiae*. *Genes Dev.* **18,** 1439–1451 (2004).
26. King, E. M., van der Sar, S. J. & Hardwick, K. G. Mad3 KEN boxes mediate both Cdc20 and Mad3 turnover, and are critical for the spindle checkpoint. *PLoS One* **2,** e342 (2007).
27. Ge, S., Skaar, J. R. & Pagano, M. APC/C- and Mad2-mediated degradation of Cdc20 during spindle checkpoint activation. *Cell Cycle* **8,** 167–171 (2009).
28. Foster, S. A. & Morgan, D. O. The APC/C subunit Mnd2/Apc15 promotes Cdc20 autoubiquitination and spindle assembly checkpoint inactivation. *Mol. Cell* **47,** 921–932 (2012).
29. Uzunova, K. *et al.* APC15 mediates CDC20 autoubiquitylation by APC/C(MCC) and disassembly of the mitotic checkpoint complex. *Nat. Struct. Mol. Biol.* **19,** 1116–1123 (2012).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

30. Mansfeld, J., Collin, P., Collins, M. O., Choudhary, J. S. & Pines, J. APC15 drives the turnover of MCC-CDC20 to make the spindle assembly checkpoint responsive to kinetochore attachment. *Nat. Cell Biol.* **13,** 1234–1243 (2011).
31. Herzog, F. *et al.* Structure of the anaphase-promoting complex/cyclosome interacting with a mitotic checkpoint complex. *Science* **323,** 1477–1481 (2009).
32. Chang, L., Zhang, Z., Yang, J., McLaughlin, S. H. & Barford, D. Atomic structure of the APC/C and its mechanism of protein ubiquitination. *Nature* **522,** 450–454 (2015).
33. Malureanu, L. A. *et al.* BubR1 N terminus acts as a soluble inhibitor of cyclin B degradation by APC/C(Cdc20) in interphase. *Dev. Cell* **16,** 118–131 (2009).
34. Suijkerbuijk, S. J. *et al.* The vertebrate mitotic checkpoint protein BUBR1 is an unusual pseudokinase. *Dev. Cell* **22,** 1321–1329 (2012).
35. Han, J. S. *et al.* Catalytic assembly of the mitotic checkpoint inhibitor BubR1-Cdc20 by a Mad2-induced functional switch in Cdc20. *Mol. Cell* **51,** 92–104 (2013).
36. Lara-Gonzalez, P., Scott, M. I., Diez, M., Sen, O. & Taylor, S. S. BubR1 blocks substrate recruitment to the APC/C in a KEN-box-dependent manner. *J. Cell Sci.* **124,** 4332–4345 (2011).
37. Izawa, D. & Pines, J. How APC/C-Cdc20 changes its substrate specificity in mitosis. *Nat. Cell Biol.* **13,** 223–233 (2011).
38. He, J. *et al.* Insights into degron recognition by APC/C coactivators from the structure of an Acm1-Cdh1 complex. *Mol. Cell* **50,** 649–660 (2013).
39. Elowe, S. *et al.* Uncoupling of the spindle-checkpoint and chromosome-congression functions of BubR1. *J. Cell Sci.* **123,** 84–94 (2010).
40. Lischetti, T., Zhang, G., Sedgwick, G. G., Bolanos-Garcia, V. M. & Nilsson, J. The internal Cdc20 binding site in BubR1 facilitates both spindle assembly checkpoint signalling and silencing. *Nat. Commun.* **5,** 5563 (2014).
41. Di Fiore, B. *et al.* The ABBA motif binds APC/C activators and is shared by APC/C substrates and regulators. *Dev. Cell* **32,** 358–372 (2015).
42. Diaz-Martinez, L. A. *et al.* The Cdc20-binding Phe box of the spindle checkpoint protein BubR1 maintains the mitotic checkpoint complex during mitosis. *J. Biol. Chem.* **290,** 2431–2443 (2015).
43. Kaisari, S., Sitry-Shevah, D., Miniowitz-Shemtov, S. & Hershko, A. Intermediates in the assembly of mitotic checkpoint complexes and their role in the regulation of the anaphase-promoting complex. *Proc. Natl Acad. Sci. USA* **113,** 966–971 (2016).
44. Brown, N. G. *et al.* RING E3 mechanism for ubiquitin ligation to a disordered substrate visualized for human anaphase-promoting complex. *Proc. Natl Acad. Sci. USA* **112,** 5272–5279 (2015).
45. Varetti, G., Guida, C., Santaguida, S., Chiroli, E. & Musacchio, A. Homeostatic control of mitotic arrest. *Mol. Cell* **44,** 710–720 (2011).
46. Garnett, M. J. *et al.* UBE2S elongates ubiquitin chains on APC/C substrates to promote mitotic exit. *Nat. Cell Biol.* **11,** 1363–1369 (2009).
47. Williamson, A. *et al.* Identification of a physiological E2 module for the human anaphase-promoting complex. *Proc. Natl Acad. Sci. USA* **106,** 18213–18218 (2009).
48. Kelly, A., Wickliffe, K. E., Song, L., Fedrigo, I. & Rape, M. Ubiquitin chain elongation requires E3-dependent tracking of the emerging conjugate. *Mol. Cell* **56,** 232–245 (2014).
49. Reis, A., Levasseur, M., Chang, H. Y., Elliott, D. J. & Jones, K. T. The CRY box: a second APCcdh1-dependent degron in mammalian cdc20. *EMBO Rep.* **7,** 1040–1045 (2006).
50. Plechanovová, A., Jaffray, E. G., Tatham, M. H., Naismith, J. H. & Hay, R. T. Structure of a RING E3 ligase and ubiquitin-loaded E2 primed for catalysis. *Nature* **489,** 115–120 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by a Cancer Research UK grant (C576/A14109) and the Medical Research Council (MRC_UP_1201/6) to D.B. and a Long Term EMBO Fellowship to C.A. We thank members of the Barford group for discussions, X. Bai and S. Scheres for their help with RELION; C. Savva and S. Chen for EM facilities; J. Grimmett and T. Darling for computing and W. Zachariae and J. Pines for their invaluable advice and J. Pines for communicating data before publication.

Author Contributions Z.Z. cloned an initial mutant MCC construct used by C.A. to generate other MCC constructs. C.A. cloned Cdc20, purified proteins, performed the protein complex reconstitutions and biochemical analysis. C.A. and L.C. prepared grids, collected and analysed EM data and determined the 3D reconstructions. C.A. fitted coordinates, built models and made the figures with help of L.C. Z.Z. together with J.Y. generated APC/C viruses. J.Y. provided E1, E2 and ubiquitin for ubiquitination assays. S.M. and M.S. performed mass spectrometry. D.B. directed the project and designed experiments with C.A. C.A. and D.B. wrote the manuscript.

Author Information EM maps are deposited with the Electron Microscopy Data Bank (EMDB) under accession code EMD-4037. Protein coordinates of APC/C^{MCC} are deposited with the Protein Data Bank (PDB) under accession code 5LCW. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.B. (dbarford@mrc-lmb.cam.ac.uk).

Reviewer Information *Nature* thanks A. Leschziner and the other anonymous reviewer(s) for their contribution to the peer review of this work.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Cloning and expression of human APC/C complex genes. The DNA coding sequences (CDSs) of the human APC/C subunits (wild-type, mutant Apc2$^{\Delta WHB}$, Apc11–UbcH10 fusion and ΔApc15) were assembled by USER cloning into a modified version of the insect cell-baculovirus MultiBac expression system[32,51,52]. All APC/C subunit CDSs were distributed in two recombinant vectors that were used for recombinant baculovirus generation. For APC/C expression, Hi-5 cells at a density of 2×10^6 cells ml^{-1} were co-infected with two pre-cultures of Sf9 cells each pre-infected with one of the two recombinant APC/C baculoviruses. APC/C expression (unphosphorylated) was performed for 30 h. To obtain APC/C^{OA} (phosphorylated APC/C), okadaic acid at a final concentration of 0.1 μM was added after 24 h of infection. Cells were collected after 5 h of treatment.

Cloning and expression of human MCC complex genes and Cdc20. The CDSs of the human MCC subunits (Mad2, Cdc20, BubR1 and Bub3) used for structural analysis were cloned into a pU2 plasmid[52] using the same method as for the APC/C. BubR1 was fused in frame with an N-terminal 3×Flag tag. Cdc20 for individual expression was cloned into a pFastbac1HTA in frame with the His_6-tag. In addition, a maltose-binding protein (MBP) tag, followed by a TEV site between the starting codon of Cdc20 and the N-terminal His_6 tag, was added by restriction free cloning method (RF-cloning[53]). To obtain a vector containing Mad2, Cdc20 and BubR1 (residues 1–569) CDSs (miniMCC construct), a Mad2- and Cdc20-containing expression cassette from a pU1 vector was shuttled (by the AvrII and PmeI sites) into a pFastbacDual vector (BstZ171 and SpeI sites) that contained 3×Flag–BubR1$^{1-569}$ under the control of the p10 promoter. A C-terminal StrepIIx2 tag was added by RF-cloning into the BubR1 constructs used in ubiquitination assays. Expression of either the MCC or Cdc20 constructs was performed similarly to the APC/C (unphosphorylated) to avoid CDK-dependent inhibition of APC/C-Cdc20 interactions[54,55]. Moreover, cells were collected 48 h after infection. To express MCC complexes with the tagged versions of BubR1, virus containing the BubR1-StrepII constructs was co-infected with MCC virus. To express the MCC complex with the Cdc20^{K485R,K490R} mutations, viruses containing the individual MCC subunits were used for co-infection. Apc15$^{\Delta NTH}$, a mutant form of Apc15 with a $(Gly\text{-}Ser\text{-}Ala)_6$ linker substitution of the N-terminal helix (NTH: residues 23–57) was cloned into an *Escherichia coli* pOPIN expression vector and purified using a C-terminal StrepII^{x2} tag.

Reconstitution and purification of APC^{MCC} complexes. To generate mitotic phosphorylated APC/C (APC/C^{OA}) we incubated APC/C expressing insect cells with the phosphatase inhibitor okadaic acid (OA) (as described above). The extent of APC/C phosphorylation was monitored by assessing the migration of the Apc3 subunit on SDS–PAGE[56] (Extended Data Fig. 1a, f). The recombinant APC/C^{OA} was phosphorylated on ~110 sites (Extended Data Table 3), correlating closely with those previously identified in endogenous APC/C isolated from HeLa cells arrested by the mitotic checkpoint[56–58], and with sites phosphorylated *in vitro* by the mitotic APC/C activating kinases Cdk2-cyclinA2-Cks2 and Plk1 (ref. 22) (Extended Data Table 3). Compared with APC/C from untreated insect cells, and using Cdc20 as the coactivator, APC/C^{OA} readily ubiquitinates securin (Extended Data Fig. 1g, h).

The APC/C^{MCC} complex was reconstituted by co-lysing APC/C^{OA} expressing cells with insect cells expressing separately MBP-tagged Cdc20 and the MCC (BubR1, Bub3, Mad2 and untagged Cdc20). Hi-5 cell pellets expressing either APC/C^{OA} or MBP–Cdc20 or MCC were mixed together in reconstitution buffer containing 50 mM Hepes (pH 8.2), 150 mM NaCl, 5% glycerol, 0.5 mM TCEP, 1 mM EDTA, 0.1 mM PMSF, 2 mM benzamidine, 5 U ml^{-1} benzonase (Novagen), Complete EDTA-free protease inhibitors (Roche), 50 mM NaF, 20 mM β-glycerophosphate and 0.1 μM okadaic acid. After complete mixing the cells were co-lysed by sonication and the lysate was centrifuged for 60 min at 17,000*g*. The soluble fraction was loaded onto a Strep-Tactin Superflow Cartridge (Qiagen) for purification using the StrepII^{x2} tag on Apc4 as described previously[21]. The eluate was then applied to an anti-Flag M2 Affinity Gel (A220, Sigma) column (directed against the N-terminal Flag tag on BubR1) and incubated overnight. The APC/C^{MCC} complex was eluted with a 3×Flag peptide at a concentration of 50 μg ml^{-1}. The resulting elution was concentrated to around 1.4 mg ml^{-1} and run on a Superose 6 3.2/300 (GE Healthcare Life Sciences) gel-filtration column pre-equilibrated with gel-filtration buffer containing 20 mM HEPES (pH 8.0), 150 mM NaCl and 0.5 mM TCEP. The gel filtration was run on a ÄKTAmicro (GE Healthcare Life Sciences) with a flow rate of 50 μl min^{-1}.

An SDS–PAGE of purified APC/C^{MCC} showed both versions of Cdc20, consistent with the incorporation of two distinct subunits of Cdc20 into APC/C^{MCC} (refs 2, 20) (Extended Data Fig. 1j). Reconstituted APC/C^{MCC} is stable and homogeneous as shown by size-exclusion chromatography (Extended Data Fig. 2a).

The APC/C$^{Apc15\Delta NTH}$ complex was reconstituted by incubating recombinant APC/C$^{\Delta Apc15}$ with Apc15$^{\Delta NTH}$, at concentrations of 200 nM and 1 μM, respectively, followed by size exclusion chromatography. Anti-Apc15 antibodies were from Santa Cruz Biotechnology (sc-398448).

APC/C ubiquitination assay. To examine APC/C activity towards securin, the ubiquitination assay was performed with 60 nM of recombinant human APC/C, 150 nM UBA1, 300 nM UbcH10, 300 nM Ube2S, 20 μM ubiquitin, 2 μM securin, 5 mM ATP, 0.25 mg ml^{-1} BSA and 7 nM of recombinant human Cdc20. The ubiquitination products of securin were detected by western blot with either an anti-His antibody (631212; Clontech) or an anti-securin antibody (700791; Invitrogen).

To test the activity of a pre-assembled APC/C^{MCC} complex towards Cdc20^{MCC} (Fig. 5c), ubiquitination reactions were performed with 250 nM of recombinant human APC/C$^{Cdc20\text{-}MCC}$ and 10 μM of UbcH10 (40× excess). To test the activity of APC/C towards the Cdc20^{MCC} from individually purified wild-type and mutant MCC$^{BubR1\text{-}StrepII}$ (purification by StrepII^{x2} affinity and gel-filtration columns) ubiquitination reactions were performed with 200 nM of recombinant human APC/C^{OA}, 200 nM of recombinant human Cdc20 and either 300 or 600 nM of recombinant human MCC$^{BubR1\text{-}StrepII}$ (Fig. 5d, e). Either with a pre-assembled APC/C^{MCC} complex or with a molar excess of MCC complex over free Cdc20 and APC/C only Cdc20^{MCC} ubiquitination is promoted (data not shown)[20]. Cdc20 and the ubiquitination products of Cdc20^{MCC} were detected by western blot with an anti-Cdc20 antibody (Cdc20 H-175 sc-8358; Santa Cruz Biotechnology).

Electron microscopy. Freshly purified APC/C^{MCC} samples were analysed by negative-stain EM to check the sample quality and to obtain a low-resolution reconstruction. Micrographs were collected on a 2k×2k CCD camera fitted to a FEI Spirit electron microscope at an accelerating voltage of 120 kV, operated at a nominal magnification of 42,000 with a resulting pixel size of 2.46 Å per pixel at specimen level. Defocuses were set at approximately −2 μm. Particles were automatically selected using the autoboxer program implemented in EMAN[59]. About 150 micrographs per sample were collected yielding ~10,000 particles. After 3D classification performed with RELION[60] only the prominent best class (30–40% of total amount of particles) was used for auto-refinement and final low-resolution structure determination.

Grid preparation for both negative-stain EM and cryo-EM was performed as described previously[32,51]. Cryo-EM micrographs were collected with an FEI Tecnai Polara electron microscope at an acceleration voltage of 300 kV and Falcon III direct detector. Micrographs were taken using EPU software (FEI) at a nominal magnification of 78,000, yielding a pixel size of 1.36 Å per pixel at specimen level. A total exposure time of 1.6 s were used at a dose rate of 27 electrons per pixel. Defocus range was set at −2.0 to −4.0 μm. Movie frames were recorded as described[32].

Image processing. Image processing was performed with RELION 1.4 (ref. 60). The initial steps including motion correction, CTF estimation, particle picking and particles sorting by *Z*-score and 2D classification were performed as described[32]. Selected particles were used for a first round of 3D classification with global search and a sampling angular interval of 7.5°, using a 60 Å low-pass filtered APC/C$^{Cdh1.Emi1}$ EM map as a reference[32]. Poorly characterized 3D classes, with poorly recognizable features, were discarded at this stage and the remaining particles were refined and corrected for beam-induced particle motion using particle polishing in RELION[61]. Polished particles were used for another round of 3D classification with a local search within 15° and a smaller angular sampling interval of 3.7° (Extended Data Figs 4 and 7). The reconstruction generated from all the polished particles, low-pass filtered at 40 Å, was used as reference.

To isolate particles for the APC/C$^{MCC\text{-}closed}$ state, classes showing closed-like features for the MCC–Cdc20$^{APC/C}$ module (for example, proximity to Apc2, Apc4 and Apc10; Extended Data Fig. 4, classes 1–3) were combined and refined. The resultant map was used as reference for a subsequent 3D classification performed with a soft edge mask on the MCC–Cdc20$^{APC/C}$ module (Extended Data Fig. 4). The mask was created from a map converted from the fitted coordinates of the MCC–Cdc20 module, with three pixel extension and five pixels soft edge width. The MCC–Cdc20 module coordinates were created by fitting the MCC core coordinates and isolated Cdc20 (PDB code 4AEZ)[18], on the best MCC–Cdc20$^{APC/C}$ module density map (Extended Data Fig. 4, class 1).

To isolate particles for the APC/C$^{MCC\text{-}open}$ state, classes showing open-like features for the MCC–Cdc20 module (for example, proximity to TPR lobe and loss of contact with Apc2, Apc4 and Apc10; Extended Data Fig. 4, classes 4–5) were refined together. The obtained averaged class was used as a reference for a subsequent 3D classification performed with a larger mask (6 pixel extension and 6 pixel soft edge) created with the MCC–Cdc20$^{APC/C}$ module coordinates fitted into the corresponding density in the APC/C$^{UbcH10\text{-}MCC}$ reconstruction described below (Extended Data Figs 4 and 7).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

To obtain the APC/C$^{\Delta Apc15\text{-}MCC}$ structure, the best classes from the 3D classification with local searches step were refined together (Extended Data Fig. 7a, classes 1–3).

To isolate the particles for the APC/C$^{UbcH10\text{-}MCC}$ reconstruction, instead of performing the 3D classification with local search steps, an initial classification with a large mask (similar to APC/C$^{MCC\text{-}open}$) was performed. The latter allowed the identification of a class that features both the MCC–Cdc20 module and the UbcH10-Apc11-Apc2^{WHB}-Apc2$^{\alpha/\beta\ domain}$ assembly[32]. A large mask including the latter regions was created by fitting the MCC–Cdc20$^{APC/C}$ module coordinates and the UbcH10-Apc11-Apc2^{WHB}-Apc2$^{\alpha/\beta\ domain}$ assembly (PDB code 5A31)[32] in the preliminary APC/C$^{UbcH10\text{-}MCC}$ reconstruction. The latter mask was used for a re-classification of the initial particles and allowed the isolation of the final APC/C$^{UbcH10\text{-}MCC}$ particles (Extended Data Fig. 7c).

All resolution estimates were based on the gold standard Fourier shell correlation (FSC) = 0.143 criterion[62]. Final FSC curves were calculated using a soft mask (five pixel extension and three pixel soft edge) of the two independent reconstructions. To visualize high-resolution details, all density maps were corrected for the modulation transfer-function of the detector and sharpened by applying negative B-factors, estimated using automated procedures.

Local resolution maps for all the cryo-EM reconstructions were calculated with RESMAP[63] using a resolution range between 3.5 and 15 Å and displayed with Chimera[64]. For comparing structural features among the cryo-EM reconstructions, shown in Fig. 4 and Extended Data Fig. 3, which have different overall resolutions, a common filter of 8.5 Å was applied. This was selected based on the local resolution of the APC/C$^{UbcH10\text{-}MCC}$ map in the region assigned to Apc15 (the main region of relative comparison). APC/C$^{UbcH10\text{-}MCC}$ is the APC/C reconstruction with the lowest overall resolution. Filtering all the reconstruction to 8.5 Å resolution allowed a clear definition of the structural details of Apc15 and other regions without the appearance of noise. To visualize the connecting density between UbcH10 and Cdc20 the APC/C$^{UbcH10\text{-}MCC}$ map was filtered to 12 Å resolution based on the local resolution of this area and the threshold was slightly lowered.

Model building. Initial fitting and superposition of coordinates was performed with Chimera[64]. Model building of APC/C^{MCC} was performed in COOT[65]. APC/C platform, TPR lobe, Apc10 and accessory subunit coordinates from the atomic structure of APC/C$^{Cdh1.Emi1}$ (PDB code 4UI9)[32] were individually rigid body fit into the APC/C$^{MCC\text{-}closed}$ cryo-EM density. A few regions such as Apc4^{HBD}, Apc5^{NTD} and Apc11 were also modified by flexible fitting. The Apc2^{WHB} domain (PDB code 4YII)[44] was rigid body fit into the corresponding density. Cdc20$^{APC/C}$ IR tail and NTD were rigid body fit from the coordinates of APC/C$^{Cdc20\text{-}Hsl1}$ cryo-EM structure[22]. The Cdc20^{MCC} IR tail was modelled by superposing the TPR domain of Apc3 including Cdc20^{IR} from APC/C$^{Cdc20\text{-}Hsl1}$ to the TPR domain of APC/C^{MCC} Apc8A. Two copies of human the Cdc20^{WD40} domain (PDB code 4GGA)[66], human C-Mad2 (PDB ID: 2V64)[8] and the human BubR1^{TPR} domain (PDB code 3SI5)[67] were rigid body fit on the MCC–Cdc20 module density. Cdc20^{MCC} CRY box, included in the human Cdc20^{WD40} domain crystal structure (PDB code 4GGA)[66] was modelled by flexible fitting. In addition, the Cdc20 KILR motif was modelled by rigid body fit of the MCC core crystal structure (PDB code 4AEZ)[18] into the corresponding density. A similar procedure was applied to model the first KEN1 and helix–loop–helix region of BubR1. BubR1 D1 and D2 were modelled by rigid body fit of Acm1 D-box 3 (PDB code 3BH6)[38]. Similarly BubR1 A1 and K2 were modelled by flexible fitting of the Acm1 region spanning the A-motif and KEN box as explained in the main text. BubR1 A2 was modelled as a rigid body fit of the Acm1 A-motif. Loop extensions were modelled as idealized polyalanine.

Model refinement was performed with REFMAC 5.8 (ref. 68). A REFMAC weight of 0.04 was defined by cross-validation using half reconstructions[69]. A resolution limit of 4.0 Å was used. All available crystal structures or NMR structures were used for secondary structure restraints. The refinement statistics are summarized in Extended Data Table 2b.

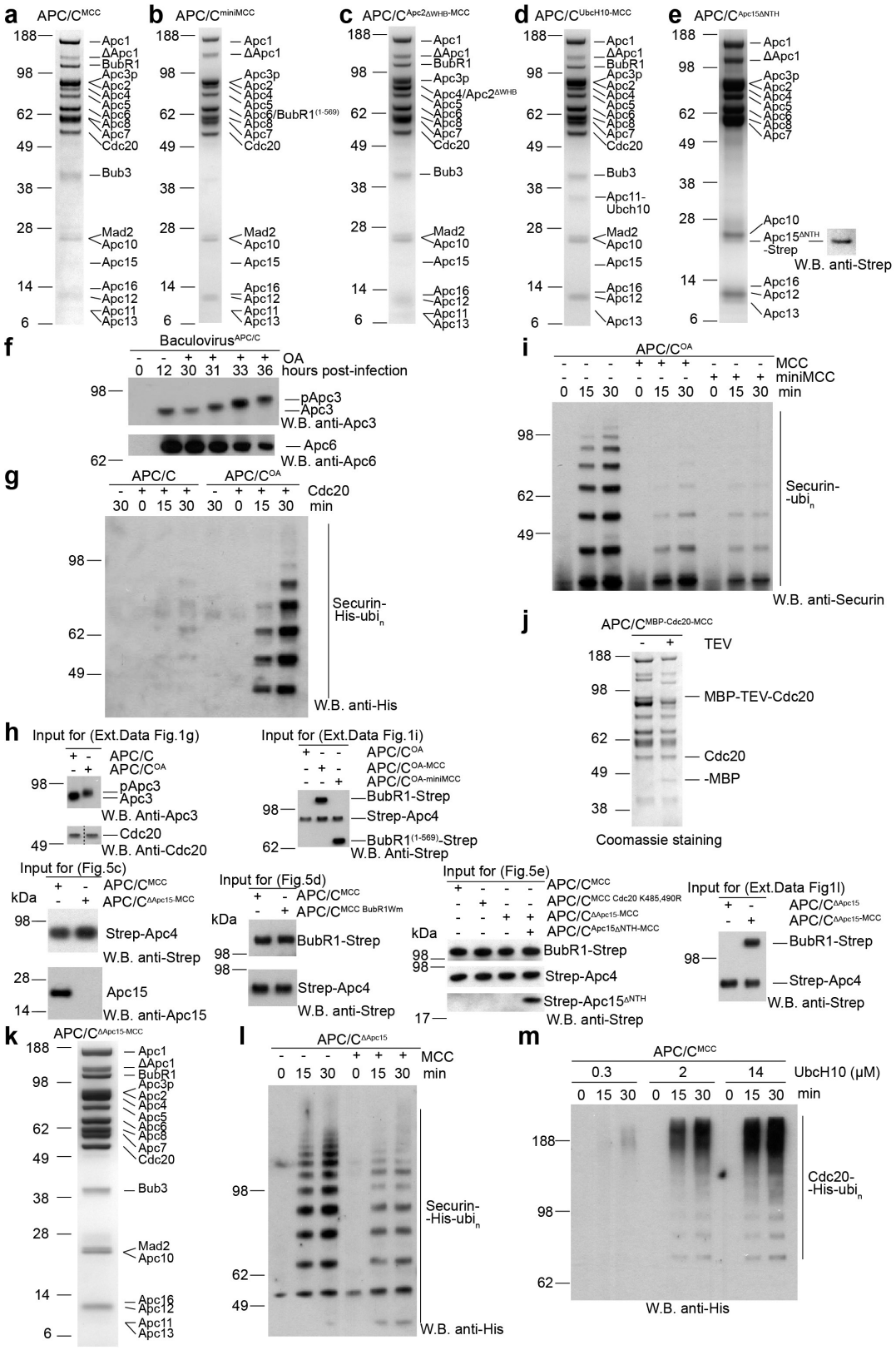
Map visualization. Figures were generated using Pymol and Chimera[70]. Structural conservation figures were generated using ConSurf[71].

Mass spectroscopy. Purified proteins were prepared for mass spectrometric analysis by in solution enzymatic digestion, without prior reduction and alkylation. Protein samples were digested with trypsin or elastase (Promega), both at an enzyme to protein ratio of 1:20. The resulting peptides were analysed by nano-scale capillary LC-MS/MS using an Ultimate U3000 HPLC (ThermoScientific Dionex) to deliver a flow of approximately 300 nl min^{-1}. A C18 Acclaim PepMap100 5 μm, 100 μm × 20 mm nanoViper (ThermoScientific Dionex), trapped the peptides before separation on a C18 Acclaim PepMap100 3 μm, 75 μm × 250 mm nanoViper (ThermoScientific Dionex). Peptides were eluted with a 90-min gradient of acetonitrile (2% to 50%). The analytical column outlet was directly interfaced via a nano-flow electrospray ionization source, with a hybrid quadrupole orbitrap mass spectrometer (Q-Exactive Plus Orbitrap, ThermoScientific). LC–MS/MS data were then searched against an in house LMB database using the Mascot search engine (Matrix Science)[72], and the peptide identifications validated using the Scaffold program (Proteome Software Inc.)[73]. All data were additionally interrogated manually.

Sequence alignment. Sequence alignment was performed using Jalview[74].

51. Zhang, Z. *et al.* Recombinant expression, reconstitution and structure of human anaphase-promoting complex (APC/C). *Biochem. J.* **449,** 365–371 (2013).
52. Zhang, Z., Yang, J. & Barford, D. Recombinant expression and reconstitution of multiprotein complexes by the USER cloning method in the insect cell-baculovirus expression system. *Methods* **95,** 13–25 (2016).
53. van den Ent, F. & Löwe, J. RF cloning: a restriction-free method for inserting target genes into plasmids. *J. Biochem. Biophys. Methods* **67,** 67–74 (2006).
54. Yudkovsky, Y., Shteinberg, M., Listovsky, T., Brandeis, M. & Hershko, A. Phosphorylation of Cdc20/fizzy negatively regulates the mammalian cyclosome/APC in the mitotic checkpoint. *Biochem. Biophys. Res. Commun.* **271,** 299–304 (2000).
55. Labit, H. *et al.* Dephosphorylation of Cdc20 is required for its C-box-dependent activation of the APC/C. *EMBO J.* **31,** 3351–3362 (2012).
56. Kraft, C. *et al.* Mitotic regulation of the human anaphase-promoting complex by phosphorylation. *EMBO J.* **22,** 6598–6609 (2003).
57. Hegemann, B. *et al.* Systematic phosphorylation analysis of human mitotic protein complexes. *Sci. Signal.* **4,** rs12 (2011).
58. Steen, J. A. *et al.* Different phosphorylation states of the anaphase promoting complex in response to antimitotic drugs: a quantitative proteomic analysis. *Proc. Natl Acad. Sci. USA* **105,** 6069–6074 (2008).
59. Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128,** 82–97 (1999).
60. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180,** 519–530 (2012).
61. Bai, X. C., Fernandez, I. S., McMullan, G. & Scheres, S. H. Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. *eLife* **2,** e00461 (2013).
62. Chen, S. *et al.* High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135,** 24–35 (2013).
63. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11,** 63–65 (2014).
64. Pettersen, E. F. *et al.* UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).
65. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 486–501 (2010).
66. Tian, W. *et al.* Structural analysis of human Cdc20 supports multisite degron recognition by APC/C. *Proc. Natl Acad. Sci. USA* **109,** 18419–18424 (2012).
67. Bolanos-Garcia, V. M. *et al.* Structure of a Blinkin-BUBR1 complex reveals an interaction crucial for kinetochore-mitotic checkpoint regulation via an unanticipated binding Site. *Structure* **19,** 1691–1700 (2011).
68. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67,** 355–367 (2011).
69. Fernández, I. S., Bai, X. C., Murshudov, G., Scheres, S. H. & Ramakrishnan, V. Initiation of translation by cricket paralysis virus IRES requires its translocation in the ribosome. *Cell* **157,** 823–831 (2014).
70. Yang, Z. *et al.* UCSF Chimera, MODELLER, and IMP: an integrated modeling system. *J. Struct. Biol.* **179,** 269–278 (2012).
71. Landau, M. *et al.* ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* **33,** W299–W302 (2005).
72. Perkins, D. N., Pappin, D. J., Creasy, D. M. & Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20,** 3551–3567 (1999).
73. Keller, A., Nesvizhskii, A. I., Kolker, E. & Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **74,** 5383–5392 (2002).
74. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. Jalview Version 2–a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25,** 1189–1191 (2009).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

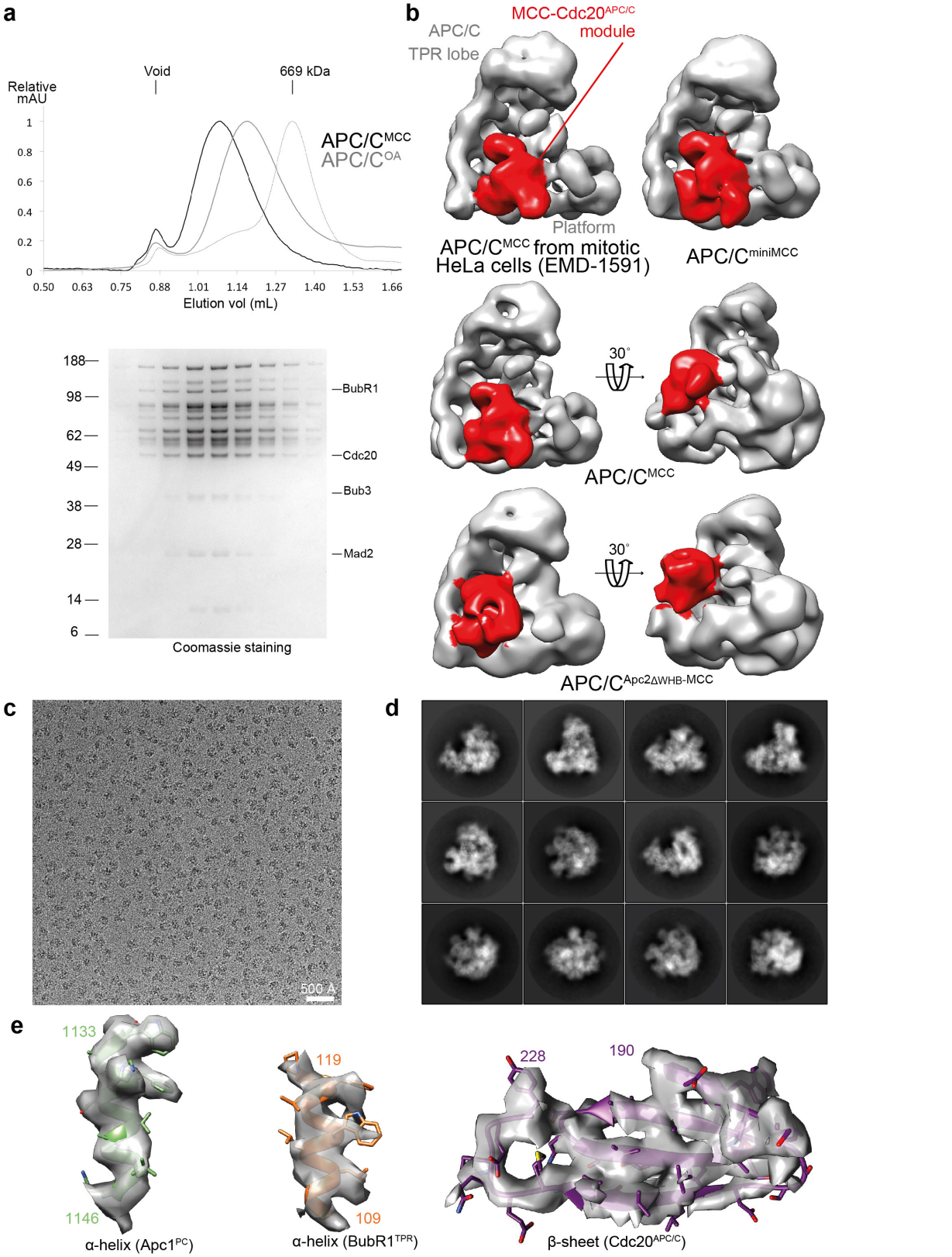


Extended Data Figure 1 | See next page for caption.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

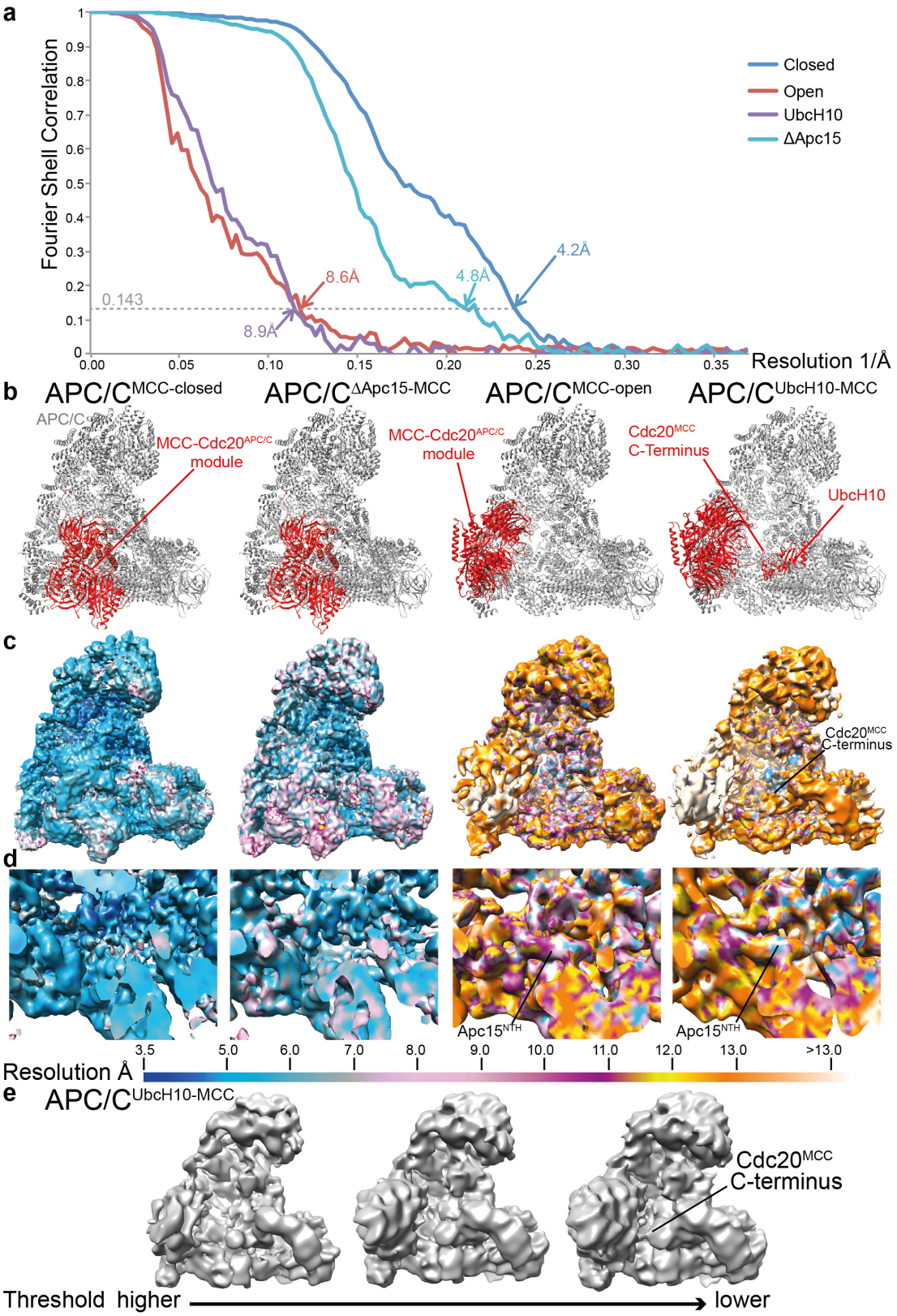
Extended Data Figure 1 | Biochemical characterization of recombinant APC/C^{MCC} complex and preparations of wild-type and mutant complexes. a–e, SDS–PAGE gels (stained with Coomassie) of gel-filtration peak fraction from wild-type and mutant APC/C^{MCC} complex preparations used in this study. Western blot against Strep tag in **e** confirms the presence of Apc15$^{\Delta NTH\text{-}Strep}$ construct. **f**, Top, western blot performed with anti-Apc3 antibody to monitor the time-dependent phosphorylation of this subunit induced by okadaic acid (OA) treatment in the APC/C-expressing insect cells. Bottom, western blot against Apc6 was used as a loading control and reflects the decrease in cell viability after addition of OA (data not shown). **g**, Western blot against His_6-tagged ubiquitin of *in vitro* securin ubiquitination assays performed with either APC/C or APC/C^{OA} in the presence or absence of Cdc20. **h**, The input sample for the ubiquitination assays performed in this study is shown. **i**, Western blot against securin of *in vitro* securin ubiquitination assays performed with APC/C^{OA} and Cdc20 with or without either MCC or miniMCC. **j**, SDS–PAGE of APC/C^{MCC} reconstituted with MBP-TEV-Cdc20$^{APC/C}$ and untagged Cdc20^{MCC}. MBP-TEV-Cdc20$^{APC/C}$ TEV cleavage products are indicated. **k**, SDS–PAGE gels of reconstituted APC/C$^{\Delta Apc15\text{-}MCC}$ complex. **l**, Western blot against His_6-tagged ubiquitin of *in vitro* securin ubiquitination assays performed with APC/C$^{\Delta Apc15}$ and Cdc20 with or without MCC. **m**, Western blot against His_6-tagged ubiquitin of *in vitro* Cdc20 ubiquitination assays performed with APC/C^{MCC} and increasing concentrations of UbcH10. Experiments in **g**, **l** and **m** were replicated three times, and those in **i** were replicated four times. See Supplementary Fig. 1 for gel source data.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



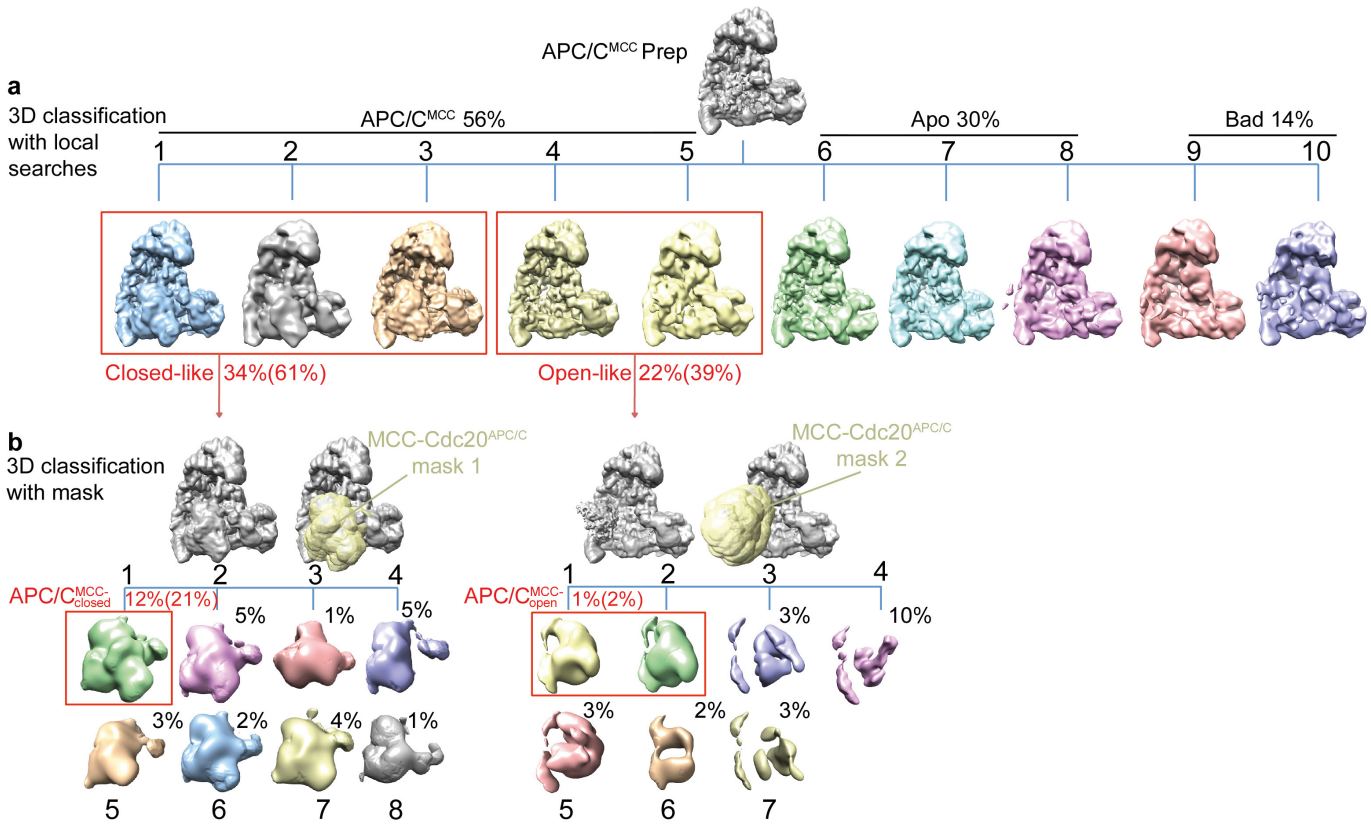
Extended Data Figure 2 | Stability of APC/C^{MCC} complex, negative-stain EM reconstructions of APC/C^{MCC} wild-type and mutant complexes and cryo-EM analysis. a, Top, chromatogram showing the elution profile of the APC/C^{MCC} complex run on a Superose 6 column. Apo APC/C^{OA} and thyroglobulin standard molecular mass marker (669 kDa) are indicated. Bottom, SDS–PAGE of the eluted fractions. APC/C^{MCC} elutes earlier than APC/C^{OA}. **b**, Negative-stain EM reconstructions performed for this study and EMD-1591 (ref. 31) are shown. APC/C (grey) and MCC–Cdc20$^{APC/C}$ module (red) are highlighted. The APC/C^{MCC} and APC/C$^{Apc2\Delta WHB\text{-}MCC}$ reconstructions are also shown in the same orientation as in Fig. 4 to facilitate comparisons. **c**, A typical cryo-EM micrograph of APC/C$^{MCC\text{-}closed}$ representative of 20,234 micrographs. **d**, Gallery of two-dimensional class averages of APC/C$^{MCC\text{-}closed}$ showing different views representative of 50 two-dimensional classes. **e**, Density quality for secondary structures. The APC/C$^{MCC\text{-}closed}$ map was filtered to 4.0 Å.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



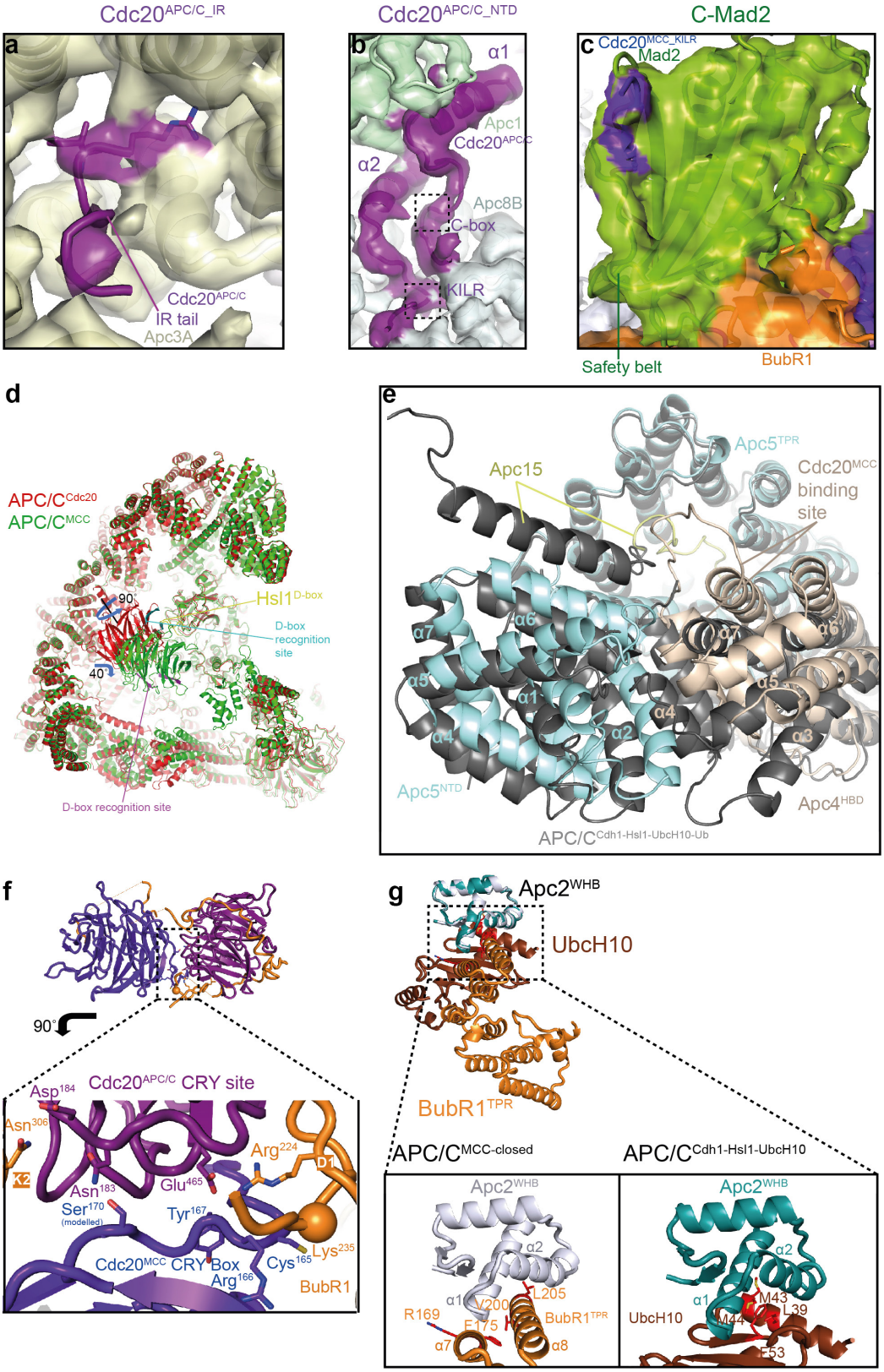
Extended Data Figure 3 | Resolution and other cryo-EM features of APC/C^{MCC} complexes. a–d, Fourier shell correlation (FSC) curves (**a**), and local resolution maps calculated with RESMAP[63] (**b–d**) are shown for all the cryo-EM reconstructions determined in this study. **b**, Ribbon representations of structures shown. **c**, Overall views of local resolution maps. **d**, Close up of platform region (Apc4, Apc5 and Apc15). All the maps shown in **c** and **d** are filtered to 8.5 Å. Local resolution colour scheme is indicated in the bar at the bottom of **d**. **e**, The APC/C$^{UbcH10\text{-}MCC}$ reconstruction filtered at 12 Å and shown at different threshold levels. The lowest threshold is the same as in Fig. 5b.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



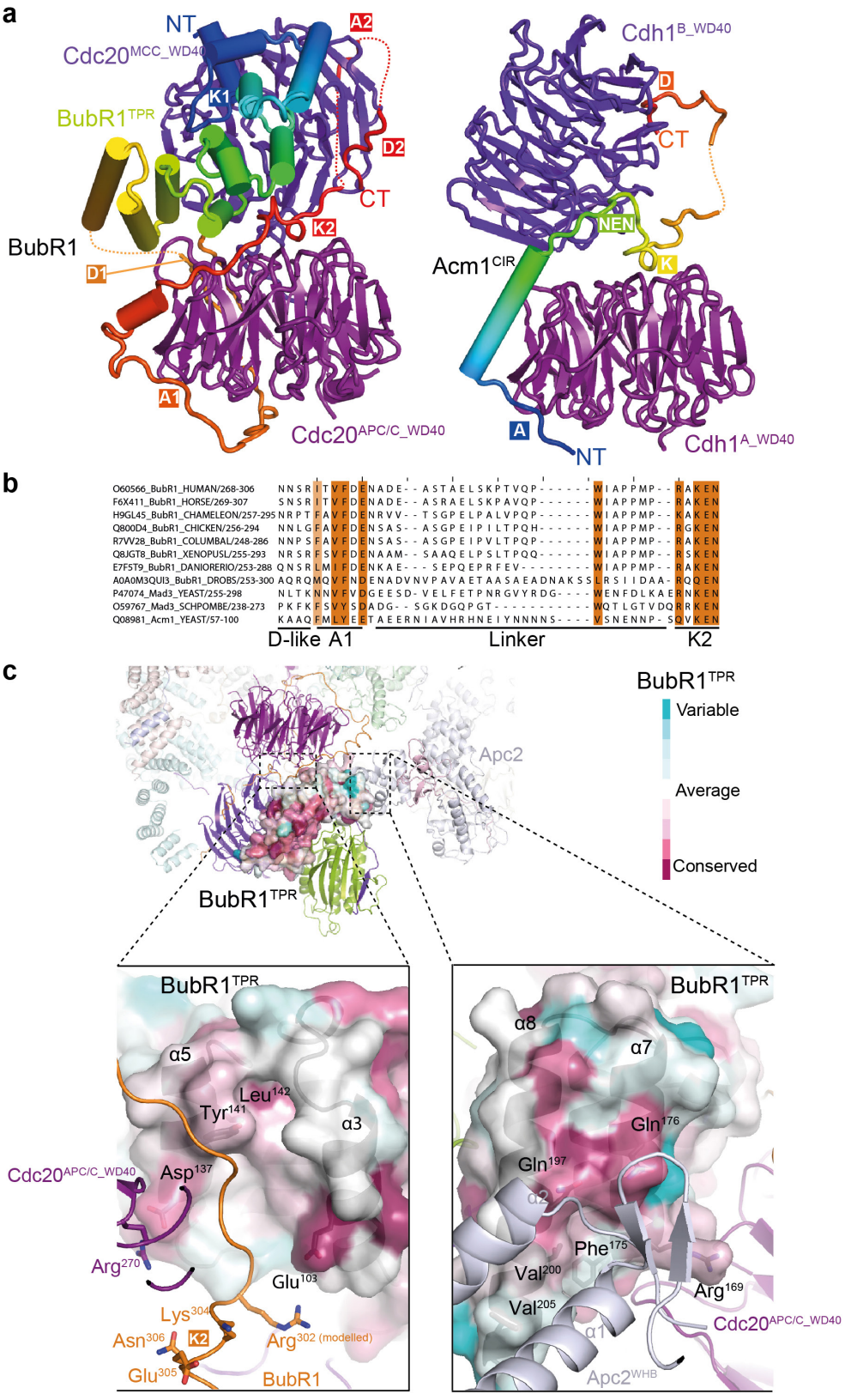
Extended Data Figure 4 | Three-dimensional classification of APC/C^{MCC}. **a**, **b**, 3D class averages obtained by classification with local searches (see Methods) are shown in **a**. Classes 1–5 (56%) show density for the MCC–Cdc20$^{APC/C}$ module with classes 1–3 and 4–5 having a closed-like and an open-like conformation, respectively (framed in red, see Methods). Particles from classes 1–3 and 4–5 were separately refined and re-classified using a mask (yellow, see Methods) shown in **b** to isolate the best quality particles for APC/C$^{MCC\text{-}closed}$ (mask 1: left) and APC/C$^{MCC\text{-}open}$ (mask 2: right). Percentages for each of the classes relative to the total number of selected particles are indicated. The percentages relative to the total number of APC/C^{MCC} particles are shown in parentheses.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



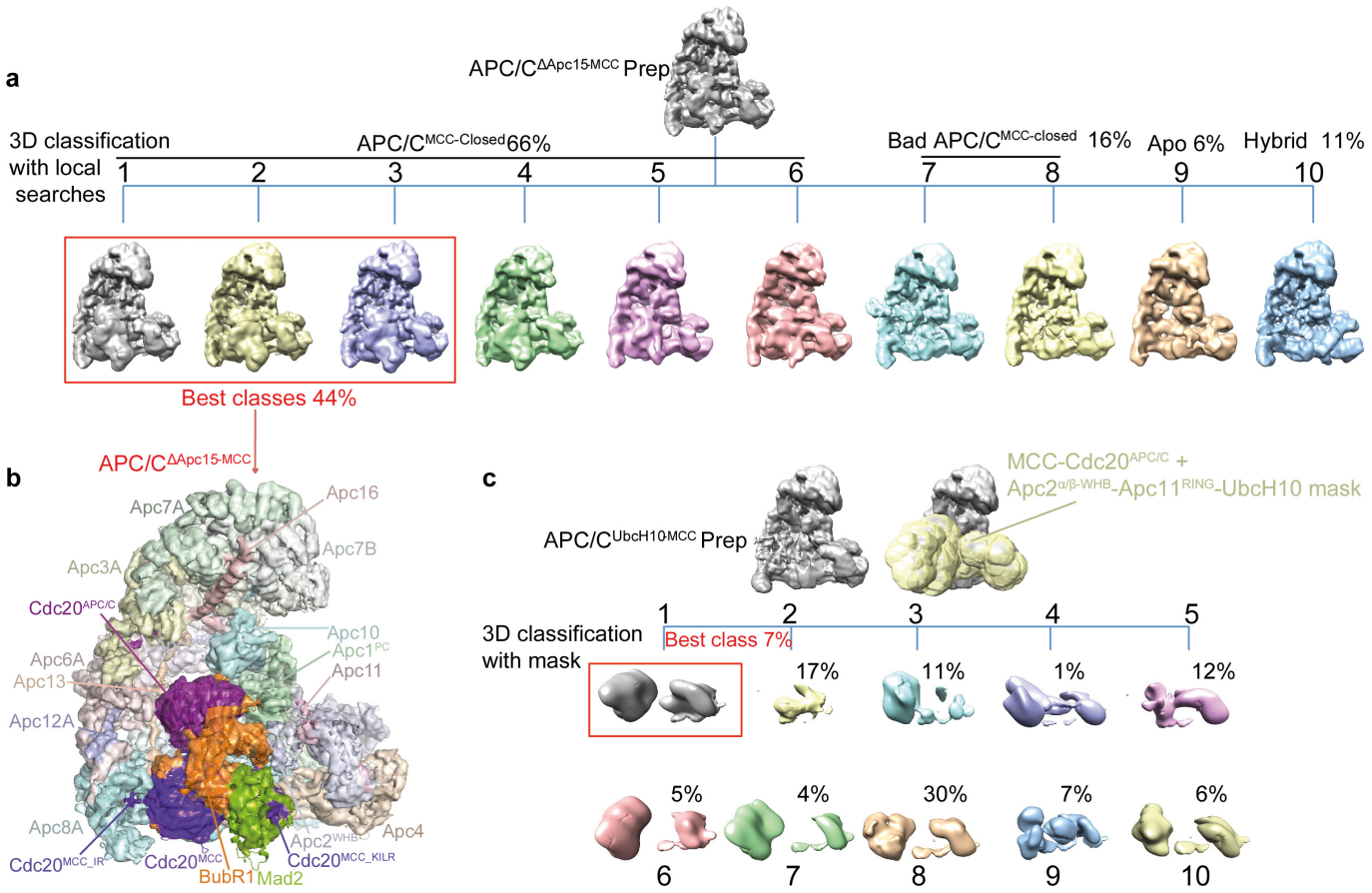
Extended Data Figure 5 | Features of the APC/C^{MCC} structure. **a**–**c**, Cryo-EM density and fitted coordinates for the Cdc20$^{APC/C}$ IR tail (Cdc20$^{APC/C_IR}$; **a**), the Cdc20$^{APC/C}$ NTD (Cdc20$^{APC/C_NTD}$); **b**) and C-Mad2 (**c**) are shown. Colours for each subunit are as in Fig. 1. **d**, Overall superposition of the APC/C$^{Cdc20\text{-}Hsl1}$ structure (red) with the APC/C$^{MCC\text{-}closed}$ structure (green). The Cdc20^{WD40} change of position is illustrated, and the blades forming the D-box (yellow) binding-pocket are highlighted. **e**, Superposition of Apc4, Apc5 and Apc15 between APC/C$^{Cdh1\text{-}Hsl1\text{-}UbcH10\text{-}Ub}$ structure (grey) and APC/C^{MCC} (subunit colours as in Fig. 1) shows the marked conformational change of Apc4^{HBD}, Apc5^{NTD} and Apc15^{NTH} induced by Cdc20^{MCC} binding to its indicated binding site on Apc4^{HBD}. **f**, Close up view of the Cdc20^{MCC} CRY box recognition site of Cdc20$^{APC/C}$. The CRY box also contacts BubR1 in proximity to D1. Colours for each subunit are as in Fig. 1. **g**, Superposition of the Apc2^{WHB} domains from APC/C^{MCC} and APC/C$^{Cdh1\text{-}Hsl1\text{-}UbcH10\text{-}Ub}$ structures and the corresponding interacting regions of BubR1^{TPR} and UbcH10 are shown. Bottom left, the residues mutated in BubR1^{Wm} that contact Apc2^{WHB} and used in the ubiquitination assay shown in Fig. 5d are indicated (red). Bottom right, residues of UbcH10 (red) that contact the corresponding site on Apc2^{WHB} ablate APC/C UbcH10-dependent ubiquitination activity[44].

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 6 | Conservation analysis on BubR1$^{A1\text{-}K2}$ and BubR1^{TPR} regions. a, Similarities in modes of binding of BubR1 to two Cdc20 subunits of APC/C^{MCC} (left) and Acm1 to two Cdh1 subunits in the Acm1–Cdh1 heterotrimer[38] (right). D-box, KEN box, NEN box and ABBA motif are labelled as D, K, NEN and A, respectively. BubR1 (colour-ramped from blue to red indicating N to C terminus) mediates a Cdc20 dimer interface, whereas Acm1 mediates a Cdh1 dimer interface. **b**, Local sequence alignment performed with BubR1$^{A1\text{-}K2}$ region sequences from several species (described on the left as: sequence identifier_protein name_species/residue number) and the *Saccharomyces cerevisiae* Acm1$^{A\text{-}KEN}$ region. A D-box-like feature (corresponding to Emi1$^{D\text{-}box}$ 7–10 positions)[32] precedes the first ABBA motif (A1). A 21–33-residue long linker connects the A1 to the second KEN-box (K2). Conserved positions are highlighted in orange. **c**, ConSurf analysis of the BubR1^{TPR} region highlighting conserved residues on the Cdc20$^{APC/C}$ binding pocket (left) and on the Apc2^{WHB} domain pocket (right). The Cdc20$^{APC/C}$ binding pocket is required for a functional SAC[67]. This site interacts with residues of BubR1 immediately N-terminal to KEN-2, thereby reinforcing their contacts with Cdc20$^{APC/C}$. Residue conservation is indicated in a gradient from cyan to purple. BubR1, Cdc20$^{APC/C}$ and Apc2^{WHB} are coloured as in Fig. 1.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 7 | Three-dimensional classification of APC/C$^{\Delta Apc15\text{-}MCC}$ and APC/C$^{UbcH10\text{-}MCC}$. **a**, 3D class averages obtained by classification with local searches (see Methods) are shown for APC/C$^{\Delta Apc15\text{-}MCC}$. Particles from classes 1–3 were refined together for obtaining the final APC/C$^{\Delta Apc15\text{-}MCC}$ reconstruction shown in **b**. The APC/C$^{\Delta Apc15\text{-}MCC}$ map was filtered to 4.8 Å. **c**, 3D class averages obtained by classification using a mask (yellow, see Methods) are shown for APC/C$^{UbcH10\text{-}MCC}$. Class 1 was used for the final reconstruction. Percentages relative to the total amount of particles are indicated for each of the classes.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 1 | Glossary of terms and abbreviations used in this study

Abbreviation	Full name
APC/C	Anaphase-promoting complex/cyclosome (subunits: Apc1 to Apc8, Apc10 to Apc13, Apc15, Apc16)
MCC	Mitotic checkpoint complex (subunits: $Cdc20^{MCC}$, Mad2, BubR1, Bub3)
SAC	Spindle assembly checkpoint
Cdc20	Mitotic APC/C coactivator subunit
Cdh1	Late mitotic APC/C coactivator subunit
UbcH10	Initiating E2
Ube2S	Elongating E2
APC/C^{Cdc20}	APC/C – Cdc20 complex
APC/C^{Cdh1}	APC/C – Cdh1 complex
APC/C^{MCC}	Inhibited APC/C^{Cdc20} – MCC complex
$APC/C^{MCC\text{-}closed}$	Closed conformation of APC/C^{MCC}
$APC/C^{MCC\text{-}open}$	Open conformation of APC/C^{MCC}
$APC/C^{UbcH10\text{-}MCC}$	APC/C^{MCC} – UbcH10 complex
$APC/C^{Cdh1\text{-}Emi1}$	APC/C^{Cdh1} – Emi1 complex
$APC/C^{Cdh1.substrate.UbcH10\text{-}Ub}$	APC/C^{Cdh1} – substrate-UbcH10 complex
$APC/C^{\Delta Apc15}$	APC/C with Apc15 deleted
$APC/C^{\Delta Apc15\text{-}MCC}$	APC/C^{MCC} with Apc15 deleted
APC/C^{OA}	Reconstituted phosphorylated APC/C isolated from BV/insect cells cultures incubated with okadaic acid
$Cdc20^{APC/C}$	Cdc20 subunit of APC/C^{Cdc20}
$Cdc20^{MCC}$	Cdc20 subunit of MCC
$Cdc20^{APC/C\text{-}WD40}$	WD40 domain of Cdc20 subunit of APC/C^{Cdc20}
C-Mad2	Closed conformation of Mad2
O-Mad2	Open conformation of Mad2
IR tail	Apc3 and Apc8 interacting C-terminal Ile-Arg motif of Cdc20 (Cdh1 and Apc10)
C box	Apc8 interacting motif of Cdc20 and Cdh1 (DRYIPxR)
KILR motif	C-Mad2-interaction motif of $Cdc20^{MCC}$ and an Apc8-interacting motif of $Cdc20^{APC/C}$
LR tail	Leu-Arg-Arg-Leu C-terminal motif common to Ube2S and Emi1
D box	APC/C degron: RxxLxx[IV]xN
KEN box	APC/C degron: KEN
ABBA motif (A motif)	APC/C degron: Fx[ILV][FHY]x[DE]
CRY motif	APC/C degron: CRY
D1, D2	D-1 box of BubR1, D-2 box of BubR1
A1, A2	ABBA-1 motif of BubR1, ABBA-2 motif of BubR1
K1, K2	KEN-1 box of BubR1, KEN-2 box of BubR1
$Apc4^{HBD}$	Helix bundle domain of Apc4
$Apc5^{NTD}$	N-terminal domain domain of Apc5
$Apc15^{NTH}$	N-terminal helix of Apc15
$BubR1^{TPR}$	BubR1 tetratricopeptide (TPR) domain

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 2 | Summary of cryo-EM data and statistics

a. Statistics of all cryo-EM reconstructions

Samples	Micrographs collected	Particles used for final reconstruction	Resolution (Å)
APC/C$^{MCC\text{-}closed}$	20,234	155,263	4.2
APC/C$^{MCC\text{-}open}$		6,188	8.6
APC/C$^{UbcH10\text{-}MCC}$	2,340	8,491	8.9
APC/C$^{\Delta Apc15\text{-}MCC}$	5,054	163,308	4.8

b. Statistics of all cryo-EM structure determination

Data collection				
EM	FEI Polara, 300K eV			
Detector	FEI Falcon III			
Pixel size (Å)	1.36			
Defocus range (μm)	2.0-4.0			
Reconstruction	**APC/C$^{MCC\text{-}closed}$**	**APC/C$^{MCC\text{-}open}$**	**APC/C$^{UbcH10\text{-}MCC}$**	**APC/C$^{\Delta Apc15\text{-}MCC}$**
Software	RELION 1.4	RELION 1.4	RELION 1.4	RELION 1.4
Accuracy of rotation (degrees)	1.318	2.594	3.235	1.61
Accuracy of translations (pixels)	0.843	1.763	2.09	1.08
Final resolution (Å)	4.2	8.6	8.9	4.8
Refinement	**APC/C$^{MCC\text{-}closed}$**			
Software	RefMac 5.8			
Refmac weight	0.04			
Resoluion limit (Å)	4.0			
Residue number	9,014			
Average Fourier shell correlation	0.761			
R factor	0.2505			
Rms bind length (Å)	0.0126			
Rms bond angle (°)	1.746			
Validation				
Ramachandran plot				
Preferred	8427 (93.49%)			
Allowed	353 (3.92%)			
Outliers	234 (2.60%)			

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 3 | Summary of phosphorylation sites in APC/C complexes used in this study

Protein	Phospho-sites	untreated	OA	Reference
Apc1	LWSS51DGA			57
	QEVT65IHE			57
	PPGS202PRE			56,57,58
	RVKS276EEE			56
	QGGT291PQN			56,57,58
	NVAT297SSS			56,57,58
	HLRS307LSK			57 (Interphase)
	RSLS309KGD			57
	KGDS313PVT			57
	SPVT316SPF			57
	HSRS341PSI			56,57,58
	RSPS343ISN			56,57,58
	PSIS345NMA			56,57
	AALS351RAH			56,57,58
	RAHS355PAL			56,57,58
	GVHS362FSG			56,57,58
	HSFS364GVQ			56,57
	FNIS372SHN			56,57
	NISS373HNQ			56,57
	HNQS377PKR			56,57,58
	ISHS386PNS			57 (Interphase)
	SNGS394FLA			57
	RPST530PLD			57,58
	DGVS536TPK			56
	GVST537PKP			56,57
	LLGS547LDE			57
	VLLS555PVP			57,58
	LHDS569LYN			57
	SLY571NED			56
	GSLS688PVI			56,57,58
	ARPS699ETG			57
	ETGS703DDD			57
	LCLS731PSE			58
	RHST921SVS			57
	HSTS922VSS			57
	KHKS1347PSY			57
	AHVY1463IIA			
Apc2	ELDS205RYA			58
	RPAS314PEA			56,57,58
	GQDS470EDD			57
	RRSS499DII			
	HQFS532FSP			58
	FSFS534PER			56,58
	LIDS732DDE	N.D.		57 (Interphase)
	DDES736DSG	N.D.		57 (Interphase)
	ESDS738GMA	N.D.		57 (Interphase)

Protein	Phospho-sites	untreated	OA	Reference
Apc3	TVLT203ETP			58
	LTET205PQD			56,57,58
	PQDT209IEL			56
	SKYS225LNT			56,57,58
	AVIS241PDT			56,57
	ETPS291PGD			56
	GDGS296YLQ			56
	LQNY301TNT			56
	QNYT302NTP			56
	YTNT304PPV			56
	SVFS334QSG			58
	SGNS339REV			58
	REVT343PIL			57,58
	QTST358TPQ			58
	TSTT359PQV			58
	QVLS364PTI			56,57
	LSPT366ITS			56,58
	PTIT368SPP			56,58
	TITS369PPN			56,57,58
	LFTS384DSS			58
	SDSS387TTK			57,58
	INDS426LEI			56,57,58
	LEIT430KLD			56
	KLDS434SII			56,57
	LDSS435IIS			56,57,58
	SIIS438EGK			57,58
	STIT446PQI			56,57,58
	SSMT809DAD			
	ADDT814QLH			
	AAES821DEF			

Protein	Phospho-sites	untreated	OA	Reference
Apc4	KGKY469FNV			56
	EVLS777ESE			57,58
	LSES779EAE			56
Apc5	MELT178SRD			57
	ELTS179RDE			57
	LDVS195VRE			56,58
	KALT232PAS			58
Apc6	FEKY107LKD			56,58
	KDES112GFK			56,57
	NIIS559PPW			56,57,58
	TGLT580PLE			56,57,58
	LETS585RKT			57,58
	SRPS595LEE			56
Apc7	VRPS119TGN			56
	TGNS123AST			56,57
	NSAS125TPQ			56,57
	SAST126PQS			57,58
	DQKS545LEG			
	KEES557PTD			
Apc8	VAVT13AAV			
	QGET556PTT			56,57,58
	TPTT559EVP			56,58
	ETPT564TEV			56,57,58
	NTPT584RRV			56,57
	RRVS588PLN			56,57,58
	LNLS593SVT			57
	NLSS594VTP			57
	SSVT596P			57,58

The red shadowing shows the presence of phosphorylation sites and white indicates its absence

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

LETTER

doi:10.1038/nature19106

A terrestrial planet candidate in a temperate orbit around Proxima Centauri

Guillem Anglada-Escudé[1], Pedro J. Amado[2], John Barnes[3], Zaira M. Berdiñas[2], R. Paul Butler[4], Gavin A. L. Coleman[1], Ignacio de la Cueva[5], Stefan Dreizler[6], Michael Endl[7], Benjamin Giesers[6], Sandra V. Jeffers[6], James S. Jenkins[8], Hugh R. A. Jones[9], Marcin Kiraga[10], Martin Kürster[11], María J. López-González[2], Christopher J. Marvin[6], Nicolás Morales[2], Julien Morin[12], Richard P. Nelson[1], José L. Ortiz[2], Aviv Ofir[13], Sijme-Jan Paardekooper[1], Ansgar Reiners[6], Eloy Rodríguez[2], Cristina Rodríguez-López[2], Luis F. Sarmiento[6], John P. Strachan[1], Yiannis Tsapras[14], Mikko Tuomi[9] & Mathias Zechmeister[6]

At a distance of 1.295 parsecs[1], the red dwarf Proxima Centauri (α Centauri C, GL 551, HIP 70890 or simply Proxima) is the Sun's closest stellar neighbour and one of the best-studied low-mass stars. It has an effective temperature of only around 3,050 kelvin, a luminosity of 0.15 per cent of that of the Sun, a measured radius of 14 per cent of the radius of the Sun[2] and a mass of about 12 per cent of the mass of the Sun. Although Proxima is considered a moderately active star, its rotation period is about 83 days (ref. 3) and its quiescent activity levels and X-ray luminosity[4] are comparable to those of the Sun. Here we report observations that reveal the presence of a small planet with a minimum mass of about 1.3 Earth masses orbiting Proxima with a period of approximately 11.2 days at a semi-major-axis distance of around 0.05 astronomical units. Its equilibrium temperature is within the range where water could be liquid on its surface[5].

The results presented here consist of an analysis of previously obtained Doppler measurements (pre-2016 data) and the confirmation of a signal in a specifically designed follow-up campaign in 2016. The Doppler data come from two precision radial velocity instruments, both at the European Southern Observatory (ESO): the High Accuracy Radial velocity Planet Searcher (HARPS) and the Ultraviolet and Visual Echelle Spectrograph (UVES). HARPS is a high-resolution stabilized echelle spectrometer installed at the ESO 3.6 m telescope (La Silla Observatory, Chile), the wavelength of which is calibrated using hollow cathode lamps (ThAr). HARPS has demonstrated radial velocity measurements at approximately 1 m s^{-1} precision over timescales of years[6], including measurements of low-mass stars[7]. All of the HARPS spectra were extracted and calibrated with the standard ESO Data Reduction Software and radial velocities were measured using a least-squares template matching technique[7]. HARPS data are separated into two data sets. The first set includes all of the data obtained before 2016 by several programmes (HARPS pre-2016). The second HARPS set comes from the more recent Pale Red Dot campaign (PRD hereafter), which was designed to eliminate period ambiguities using new HARPS observations and quasi-simultaneous photometry. The HARPS PRD observations consisted of one spectrum obtained almost every night between 19 January and 31 March 2016. The UVES observations used the iodine cell technique[8] and were obtained in the framework of the UVES survey for terrestrial planets around M-class dwarfs between 2000 and 2008. The spectra were extracted using the standard procedures of the UVES survey[9] and new radial velocities were obtained using up-to-date iodine reduction codes[10]. As systematic calibration errors produce correlations among the observations for each night[11], we consolidated the Doppler measurements through nightly averages to present a simpler and more conservative signal search. This led to 72 UVES, 90 HARPS pre-2016 and 54 HARPS PRD epochs. The PRD photometric observations were obtained using the Astrograph for the South Hemisphere II telescope (ASH2 hereafter[12], with S II and Hα narrowband filters) and the Las Cumbres Observatory Global Telescope network[13] (with Johnson B and V band filters), over the same time interval and similar sampling rates as the HARPS PRD observations. Further details about each campaign and the photometry are detailed in Methods. All of the time series used in this work are available as Supplementary Data.

The search and assessment of the statistical significance (see below and Methods for more details) of the signals were performed using frequentist[14] and Bayesian[15] methods. The periodograms in Fig. 1 represent the improvement of a reference statistic as a function of trial period, with the peaks representing the most probable new signals. The improvement in the logarithm of the likelihood function $\Delta\ln L$ is the reference statistic used in the frequentist framework, and its value is then used to assess the false-alarm probability (FAP) of the detection[14]. An FAP below 1% is considered suggestive of periodic variability, and anything below 0.1% is considered to be a significant detection. In the Bayesian framework, the signals are first searched using a specialized sampling method[16] that enables the exploration of multiple local maxima of the posterior density (the result of this process is the red lines in Fig. 1), and the significance of the signals is then assessed by obtaining the ratios of evidences of the models. In a Bayesian context, the evidence B_m of a model m is the integral of its Bayesian posterior density. A more detailed description and references are provided in Methods. If the evidence ratio between two models exceeds some threshold (for example, $B_1/B_0 > 10^3$), then the model in the numerator (with one planet) is favoured against the model in the denominator (no planet).

An isolated peak at about 11.2 d was recovered when all of the night averages in the pre-2016 data sets were averaged (Fig. 1a). Despite the significance of the signal, the analysis of the pre-2016 subsets produced slightly different periods depending on the noise assumptions and which of the subsets were considered. Confirmation or refutation of this signal at 11.2 d was the main driver for the proposal of the HARPS PRD campaign. The analysis of the HARPS PRD data revealed a single significant signal at approximately the same 11.3 ± 0.1 d period (Fig. 1b), but this period coincidence alone does not prove consistency with

[1]School of Physics and Astronomy, Queen Mary University of London, 327 Mile End Road, London E1 4NS, UK. [2]Instituto de Astrofísica de Andalucía – Consejo Superior de Investigaciones Científicas, Glorieta de la Astronomía S/N, E-18008 Granada, Spain. [3]Department of Physical Sciences, Open University, Walton Hall, Milton Keynes MK7 6AA, UK. [4]Carnegie Institution of Washington, Department of Terrestrial Magnetism, 5241 Broad Branch Road NW, Washington DC 20015, USA. [5]Astroimagen, C. Abad y Lasierra, 58 Bis, 6-2, 07800 Ibiza, Spain. [6]Institut für Astrophysik, Georg-August-Universität Göttingen, Friedrich-Hund-Platz 1, 37077 Göttingen, Germany. [7]McDonald Observatory, the University of Texas at Austin, 2515 Speedway, C1400, Austin, Texas 78712, USA. [8]Departamento de Astronomía, Universidad de Chile, Camino El Observatorio 1515, Las Condes, Santiago, Chile. [9]Centre for Astrophysics Research, Science & Technology Research Institute, University of Hertfordshire, Hatfield AL10 9AB, UK. [10]Warsaw University Observatory, Aleje Ujazdowskie 4, Warszawa, Poland. [11]Max-Planck-Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany. [12]Laboratoire Univers et Particules de Montpellier, Université de Montpellier, Place E. Bataillon—CC 72, 34095 Montpellier Cédex 05, France. [13]Department of Earth and Planetary Sciences, Weizmann Institute of Science, 234 Herzl Street, Rehovot 76100, Israel. [14]Astronomisches Rechen-Institut, Mönchhofstrasse 12–14, 69120 Heidelberg, Germany.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

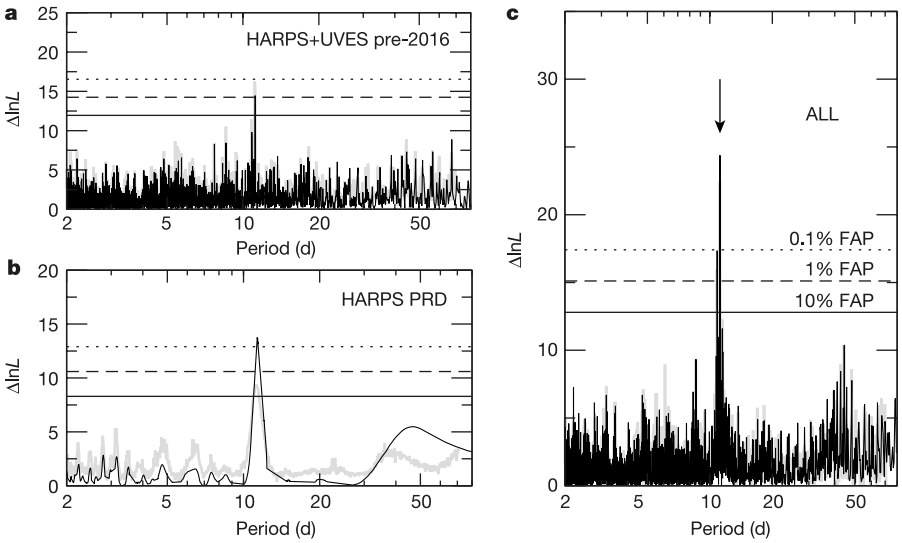


Figure 1 | Detection of a Doppler signal at 11.2 d. a, **b**, Detection periodograms of the 11.2 d signal in the HARPS+UVES pre-2016 data (**a**) and the HARPS PRD campaign only (**b**). **c**, The periodogram obtained after combining all of the data sets. Black lines correspond to the Δln*L* statistic, whereas the grey thick lines represent the logarithm of the Bayesian posterior density (see text, arbitrary vertical offset applied for visual comparison of the two statistics). The horizontal solid, dashed and dotted lines represent the FAP thresholds of the frequentist analysis.

the pre-2016 data. Final confirmation is achieved when all of the data sets are combined (Fig. 1c)—the statistical significance of the signal at 11.2 d then increases dramatically (FAP $< 10^{-7}$, Bayesian evidence ratio $B_1/B_0 > 10^6$). This implies that not only the period, but also the amplitude and phase are consistent during the 16 years of accumulated observations (see Fig. 2). All of the analyses performed with and without correlated-noise models produced consistent results. A second signal in the range of 60–500 d was also detected, but its nature is still unclear due to stellar activity and inadequate sampling.

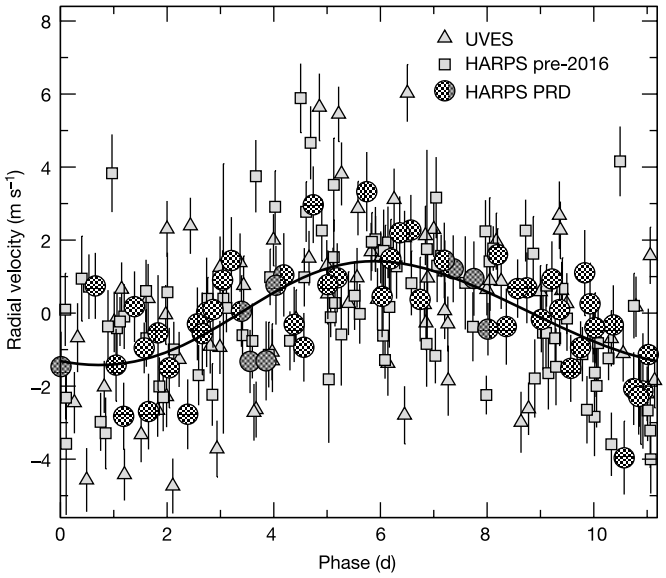


Figure 2 | All of the data sets phase-folded at the 11.2 d signal. Radial velocity measurements phase folded at the 11.2 d period of the planet candidate for 16 years of observations. The abscissa values of phase-folded plots are determined by first computing the difference between each Julian date and the reference epoch of the Keplerian fit (Julian date of the first UVES observation), and then computing the remainder of the division of this difference with the orbital period. Although its nature is unclear, a second signal at $P \approx 200$ d was fitted and subtracted from the data to produce this plot and improve visualization. Circles correspond to HARPS PRD, triangles are HARPS pre-2016 and squares are UVES. The black line represents the best Keplerian fit to this phase folded representation of the data. Error bars correspond to formal 1σ uncertainties.

Stellar variability can cause spurious Doppler signals that mimic planetary candidates, especially when combined with uneven sampling[9,17]. To address this, the time series of the photometry and spectroscopic activity indices were also searched for signals. After removing occasional flares, all four photometric time series show the same clear modulation with a period of $P \approx 80$ days (Fig. 3b–e), which is consistent with the previously reported photometric period of approximately 83 d (ref. 3). Spectroscopic activity indices were measured on all of the HARPS spectra, and their time series were also investigated. The width of the spectral lines (measured as the variance of the mean line, or m_2) follows a time dependence that is almost identical to the light curves, a behaviour that has already been reported for other M dwarf stars[18]. The time series of the indices that are based on chromospheric emission lines (for example, Hα) do not show evidence of periodic variability, even after removing the data points that are likely to be affected by flares. We also investigated possible correlations of the Doppler measurements with the activity indices by including linear correlation terms in the Bayesian model of the Doppler data. Although some indices do show hints of correlation in some campaigns, including them in the model produces lower probabilities, owing to overparameterization. Flares have very little effect on our Doppler velocities, as has already been suggested by previous observations of Proxima[19]. More details are provided in Methods and Extended Data Fig. 8. As the analysis of the activity data failed to identify any stellar activity feature that is likely to generate a spurious Doppler signal at 11.2 d, we conclude that the variability in the data is best explained by the presence of a planet (Proxima b, hereafter) orbiting the star. All of the available photometric light curves were searched for evidence of transits, but no obvious transit-like features were detectable in our light curves. We used optimal box-Least-Squares codes[20] to search for candidate signals in data from the All Sky Automatic Survey[3]. No significant transit signal was found down to a depth of about 5%. The most likely orbital solution and the putative properties of the planet and transits are given in Table 1.

The Doppler semi-amplitude of Proxima b (approximately 1.4 m s^{-1}) is not particularly small compared with other reported planet candidates[6]. The uneven and sparse sampling combined with the longer-term variability of the star seem to be the reasons why the signal could not be unambiguously confirmed with pre-2016 data rather than the total amount of data accumulated. The corresponding minimum planetary mass is about 1.3 Earth masses ($M_\oplus$). With a semi-major axis of approximately 0.05 AU, it lies squarely in the centre of the classical habitable zone for Proxima[5]. As mentioned earlier, the presence of another

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

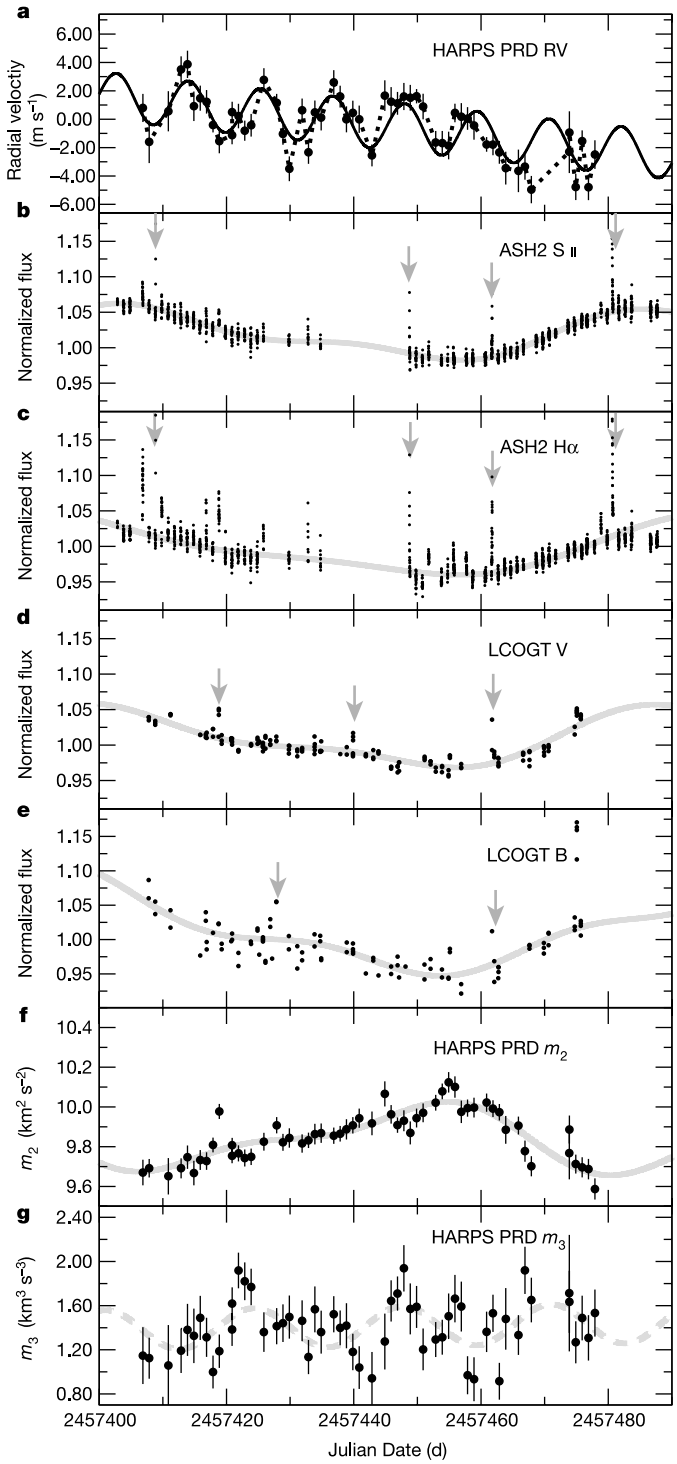


Figure 3 | Time series obtained during the PRD campaign. a, HARPS-PRD radial velocity measurements. **b**, **c**, Quasi-simultaneous photometry from ASH2 for S II (**b**) and Hα (**c**). **d**, **e**, Quasi-simultaneous photometry from LCOGT for V (**d**) and B (**e**). **f**, **g**, Central moments of the mean line profiles for m_2 (**f**) and m_3 (**g**). The solid lines show the best fits. A dashed line indicates a signal that is not sufficiently statistically significant. Excluded measurements that probably affected activity events (for example, flares) are marked with grey arrows. The photometric time series and m_2 show evidence of the same approximately 80 d modulation. Error bars correspond to formal 1σ uncertainties.

super-$M_\oplus$ planet cannot yet be ruled out at longer orbital periods and Doppler semi-amplitudes of <3 m s^{-1}. By numerical integration of some putative orbits, we verified that the presence of such a planet would not compromise the orbital stability of Proxima b.

Table 1 | Stellar properties, Keplerian parameters, and derived quantities

Stellar properties	Value	Reference
Spectral type	M5.5V	2
$M_*/M_\odot$	0.120 (0.105–0.135)	30
$R_*/R_\odot$	0.141 (0.120–0.162)	2
$L_*/L_\odot$	0.00155 (0.00149–0.00161)	2
Effective temperature (K)	3,050 (2,950–3,150)	2
Rotation period (d)	about 83	3
Habitable zone range (AU)	about 0.0423–0.0816	30
Habitable zone periods (d)	about 9.1–24.5	30
Keplerian fit	**Proxima b**	
Period (d)	11.186 (11.184–11.187)	
Doppler amplitude (m s^{-1})	1.38 (1.17–1.59)	
Eccentricity, e	<0.35	
Mean longitude, $\lambda = \omega + M_0$ (°)	110 (102–118)	
Argument of periastron, ω_0 (°)	310 (0–360)	
Statistics summary		
Frequentist FAP	7×10^{-8}	
Bayesian odds in favour, B_1/B_0	2.1×10^7	
UVES jitter (m s^{-1})	1.69 (1.22–2.33)	
HARPS pre-2016 jitter (m s^{-1})	1.76 (1.22–2.36)	
HARPS PRD jitter (m s^{-1})	1.14 (0.57–1.84)	
Derived quantities		
Orbital semi-major axis, a (AU)	0.0485 (0.0434–0.0526)	
Minimum mass, $m_p \sin i$ ($M_\oplus$)	1.27 (1.10–1.46)	
Equilibrium black body temperature (K)	234 (220–240)	
Irradiance compared with Earth	65%	
Geometric probability of transit	about 1.5%	
Transit depth (Earth-like density)	about 0.5%	

The estimates are the maximum *a posteriori* values and the uncertainties of the parameters are expressed as 68% credibility intervals. We provide only an upper limit for the eccentricity (95% confidence level). Extended Data Table 1 contains the list of all of the model parameters.

The habitability of planets like Proxima b—in the sense of sustaining an atmosphere and liquid water on its surface—is a matter of intense debate. The most common arguments against habitability are tidal locking, strong stellar magnetic fields, strong flares and high ultraviolet and X-ray fluxes; but none of these have been proved definitive. Tidal locking does not preclude a stable atmosphere via global atmospheric circulation and heat redistribution[21]. The average global magnetic flux density of Proxima is 600 ± 150 G (ref. 22), which is quite large compared with that of the Sun (1 G). However, several studies have shown that planetary magnetic fields in tidally locked planets can be strong enough to prevent atmospheric erosion by stellar magnetic fields[23] and flares[24]. Because of its close orbit to Proxima, Proxima b suffers from X-ray fluxes that are approximately 400 times that experienced by Earth, but studies of similar systems indicate that atmospheric losses can be relatively small[25]. Further characterization of such planets can also inform us about the origin and evolution of terrestrial planets. For example, the formation of Proxima b from *in situ* disk material is implausible because disk models for small stars would contain less than $1M_\oplus$ of solids within a distance of 1 AU. There are three possibilities: the planet migrated in via type I migration[26]; planetary embryos migrated in and coalesced at the current planet's orbit; or pebbles/small planetesimals migrated via aerodynamic drag[27] and later coagulated into a larger body. Although migrated planets and embryos that originate beyond the ice-line would be rich in volatiles, pebble migration would produce much drier worlds. A warm terrestrial planet orbiting Proxima offers the opportunity to attempt further characterization via transits (ongoing searches), by direct imaging and high-resolution spectroscopy in the next decades[28], and possibly robotic exploration in the coming centuries[29].

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 6 May; accepted 7 July 2016.

1. van Leeuwen, F. Validation of the new Hipparcos reduction. *Astron. Astrophys.* **474,** 653–664 (2007).
2. Boyajian, T. S. *et al.* Stellar diameters and temperatures. II: Main-sequence K- and M-stars. *Astrophys. J.* **757,** 112 (2012).
3. Kiraga, M. & Stepien, K. Age–rotation–activity relations for M dwarf stars. *Acta Astron.* **57,** 149–172 (2007).
4. Güdel, M., Audard, M., Reale, F., Skinner, S. L. & Linsky, J. L. Flares from small to large: X-ray spectroscopy of Proxima Centauri with XMM-Newton. *Astron. Astrophys.* **416,** 713–732 (2004).
5. Kopparapu, R. K. *et al.* Habitable zones around main-sequence stars: new estimates. *Astrophys. J.* **765,** 131 (2013).
6. Pepe, F. *et al.* The HARPS search for Earth-like planets in the habitable zone. I. Very low-mass planets around HD 20794, HD 85512, and HD 192310. *Astron. Astrophys.* **534,** A58 (2011).
7. Anglada-Escudé, G. & Butler, R. P. The HARPS-TERRA Project. I: description of the algorithms, performance, and new measurements on a few remarkable stars observed by HARPS. *Astrophys. J. Suppl. Ser.* **200,** 15 (2012).
8. Butler, R. P. *et al.* Attaining Doppler precision of 3 m s^{-1}. *Publ. Astron. Soc. Pacif.* **108,** 500–509 (1996).
9. Kürster, M. *et al.* The low-level radial velocity variability in Barnard's star (= GJ 699): secular acceleration, indications for convective redshift, and planet mass limits. *Astron. Astrophys.* **403,** 1077–1087 (2003).
10. Arriagada, P. *et al.* Two planetary companions around the K7 dwarf GJ 221: a hot super-Earth and a candidate in the sub-Saturn desert range. *Astrophys. J.* **771,** 42 (2013).
11. Berdiñas, Z. M., Amado, P. J., Anglada-Escudé, G., Rodríguez-López, C. & Barnes, J. High-cadence spectroscopy of M-dwarfs. I: analysis of systematic effects in HARPS-N line profile measurements on the bright binary GJ 725A+B. *Mon. Not. R. Astron. Soc.* **459,** 3551B (2016).
12. Sicardy, B. *et al.* A Pluto-like radius and a high albedo for the dwarf planet Eris from an occultation. *Nature* **478,** 493–496 (2011).
13. Brown, T. M. *et al.* Las Cumbres Observatory Global Telescope Network. *Publ. Astron. Soc. Pacific* **125,** 1031–1055 (2013).
14. Baluev, R. V. The impact of red noise in radial velocity planet searches: only three planets orbiting GJ 581? *Mon. Not. R. Astron. Soc.* **429,** 2052–2068 (2013).
15. Tuomi, M., Jones, H. R. A., Barnes, J. R., Anglada-Escudé, G. & Jenkins, J. S. Bayesian search for low-mass planets around nearby M dwarfs—estimates for occurrence rate based on global detectability statistics. *Mon. Not. R. Astron. Soc.* **441,** 1545–1569 (2014).
16. Haario, H., Laine, M., Mira, A. & Saksman, E. Dram: efficient adaptive MCMC. *Stat. Comput.* **16,** 339–354 (2006).
17. Rajpaul, V., Aigrain, S. & Roberts, S. Ghost in the time series: no planet for Alpha Cen B. *Mon. Not. R. Astron. Soc.* **456,** L6–L10 (2016).
18. Bonfils, X. *et al.* The HARPS search for southern extra-solar planets. X: A $m\sin i = 11 M_{\oplus}$ planet around the nearby spotted M dwarf GJ 674. *Astron. Astrophys.* **474,** 293–299 (2007).
19. Barnes, J. R. *et al.* Precision radial velocities of 15 M5–M9 dwarfs. *Mon. Not. R. Astron. Soc.* **439,** 3094–3113 (2014).
20. Ofir, A. Optimizing the search for transiting planets in long time series. *Astron. Astrophys.* **561,** A138 (2014).
21. Kopparapu, R. k. *et al.* The inner edge of the habitable zone for synchronously rotating planets around low-mass stars using general circulation models. *Astrophys. J.* **819,** 84 (2016).
22. Reiners, A. & Basri, G. The moderate magnetic field of the flare star Proxima Centauri. *Astron. Astrophys.* **489,** L45–L48 (2008).
23. Vidotto, A. A. *et al.* Effects of M dwarf magnetic fields on potentially habitable planets. *Astron. Astrophys.* **557,** A67 (2013).
24. Zuluaga, J. I., Bustamante, S., Cuartas, P. A. & Hoyos, J. H. The influence of thermal evolution in the magnetic protection of terrestrial planets. *Astrophys. J.* **770,** 23 (2013).
25. Bolmont, E. *et al.* Water loss from Earth-sized planets in the habitable zones of ultracool dwarfs: implications for the planets of TRAPPIST-1. Preprint at http://arxiv.org/abs/1605.00616 (2016).
26. Tanaka, H., Takeuchi, T. & Ward, W. R. Three-dimensional interaction between a planet and an isothermal gaseous disk. I: corotation and Lindblad torques and planet migration. *Astrophys. J.* **565,** 1257–1274 (2002).
27. Weidenschilling, S. J. Aerodynamics of solid bodies in the solar nebula. *Mon. Not. R. Astron. Soc.* **180,** 57–70 (1977).
28. Snellen, I. *et al.* Combining high-dispersion spectroscopy with high contrast imaging: probing rocky planets around our nearest neighbors. *Astron. Astrophys.* **576,** A59 (2015).
29. Lubin, P. A roadmap to interstellar flight. Preprint at http://arxiv.org/abs/1604.01356 (2016).
30. Delfosse, X. *et al.* Accurate masses of very low mass stars. IV. Improved mass-luminosity relations. *Astron. Astrophys.* **364,** 217–224 (2000).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank E. Gerlach, R. Street and U. Seemann for their support to the science preparations. We thank P. Micakovic, M. M. Mutter (QMUL), R. Ivison, G. Hussain, I. Saviane, O. Sandu, L. L. Christensen, R. Hook and the personnel at La Silla (ESO) for making the PRD campaign possible. The authors acknowledge support from the following funding grants: Leverhulme Trust/UK RPG-2014-281 (H.R.A.J., G.A.-E. and M.T.); MINECO/Spain AYA-2014-54348-C3-1-R (P.J.A., C.R.-L., Z.M.B. and E.R.); MINECO/Spain ESP2014-54362-P (M.J.L.-G.); MINECO/Spain AYA-2014-56637-C2-1-P (J.L.O. and N.M.); J.A./Spain 2012-FQM1776 (J.L.O. and N.M.); CATA-Basal/Chile PB06 Conicyt (J.S.J.); Fondecyt/Chile project #1161218 (J.S.J.); STFC/UK ST/M001008/1 (R.P.N., G.A.L.C. and G.A.-E.); STFC/UK ST/L000776/1 (J.B.); ERC/EU Starting Grant #279347 (A.R., L.F.S. and S.V.J.); DFG/Germany Research Grants RE 1664/9-2 (A.R.); RE 1664/12-1 (M.Z.); DFG/Germany Colloborative Research Center 963 (C.J.M. and S.D.); DFG/Germany Research Training Group 1351 (L.F.S.); and NSF/USA grant AST-1313075 (M.E.). Study based on observations made with ESO Telescopes at the La Silla Paranal Observatory under programmes 096.C-0082 and 191.C-0505. Observations were obtained with ASH2, which is supported by the Instituto de Astrofísica de Andalucía and Astroimagen. This work makes use of observations from the LCOGT network. We acknowledge the effort of the UVES/M-dwarf and the HARPS/Geneva teams, who obtained a substantial amount of the data used in this work.

Author Contributions In the author list, after G.A.-E., the authors are listed in alphabetical order. G.A.-E. led the PRD campaign, observing proposals and organized the manuscript. P.J.A. led observing proposals and organized and supported the Instituto de Astrofisica de Andalucía team through research grants. M.T. obtained the early signal detections and most of the Bayesian analyses. J.S.J., J.B., Z.M.B. and H.R.A.J. participated in the analyses and obtained activity measurements. Z.M.B. also led observing proposals. H.R.A.J. funded several co-authors via research grants. M. Kuerster and M.E. provided the extracted UVES spectra, and R.P.B. re-derived radial velocity measurements. C.R.-L. coordinated photometric follow-up campaigns. E.R. led the ASH2 team and related reductions (M.J.L.-G., I.d.l.C., J.L.O. and N.M.). Y.T. led the LCOGT proposals, campaign and reductions. M.Z. obtained observations and performed analyses on HARPS and UVES spectra. A.O. analysed time series and transit searches. J.M., S.V.J. and A.R. analysed stellar activity data. A.R. funded several co-authors via research grants. R.P.N., G.A.L.C., S.-J.P., S.D. and B.G. did dynamical studies and studied the planet formation context. M. Kiraga provided early access to time series from the ASAS survey. C.J.M. and L.F.S. participated in the HARPS campaigns. All authors contributed to the preparation of observing proposals and the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.A.-E. (g.anglada@qmul.ac.uk).

Reviewer Information *Nature* thanks A. Hatzes and D. Queloz for their contribution to the peer review of this work.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

Statistical frameworks and tools. The analyses of time series including the radial velocities and activity indices were performed by frequentist and Bayesian methods. In all cases, the statistical significance was assessed using model comparisons by performing global multiparametric fits to the data. Here we provide a minimal overview of the methods and assumptions used throughout the Letter.

Bayesian statistical analyses. The analyses of the radial velocity data were performed by applying posterior sampling algorithms called Markov chain Monte Carlo methods. We used the adaptive Metropolis algorithm[31], which has previously been applied to such radial velocity data sets[15,32]. This algorithm is simply a generalized version of the common Metropolis–Hastings algorithm[33,34] that adapts to the posterior density based on the previous members of the chain.

The likelihood functions and posterior densities of models with periodic signals are highly multimodal (that is, they have peaks in the periodograms). For this reason, in our Bayesian signal searches we applied the delayed rejection adaptive Metropolis (DRAM) method[16], which enables efficient jumping of the chain between multiple modes by postponing the rejection of a proposed parameter vector by first attempting to find a better value in its vicinity. For every given model we performed several posterior samplings with different initial values to ensure convergence to a unique solution. When we identified two or more substantial maxima in the posterior density, we typically performed several additional samplings with initial states close to those maxima. This enabled us to evaluate their relative importance in a consistent manner. We estimated the marginal likelihoods and the corresponding Bayesian evidence ratios of different models by using a simple method[35]. A more detailed description of these methods can be found in elsewhere[36].

Doppler model and likelihood function. Assuming that the ith radial velocity measurements is $m_{i,\mathrm{INS}}$ obtained at some instant t_i from an instrument INS, the likelihood function of the observations (probability of the data given a model) is given by

$$L = \prod_{\mathrm{INS}} \prod_{i} l_{i,\mathrm{INS}} \tag{1}$$

$$l_{i,\mathrm{INS}} = \frac{1}{\sqrt{2\pi(\sigma_i^2 + \sigma_{\mathrm{INS}}^2)}} \exp\left\{ -\frac{1}{2} \frac{\epsilon_{i,\mathrm{INS}}^2}{\sigma_i^2 + \sigma_{\mathrm{INS}}^2} \right\} \tag{2}$$

$$\epsilon_{i,\mathrm{INS}} = m_{i,\mathrm{INS}} - \{\gamma_{\mathrm{INS}} + \dot{\gamma}\Delta t_i + \kappa(\Delta t_i) + \mathrm{MA}_{i,\mathrm{INS}} + A_{i,\mathrm{INS}}\} \tag{3}$$

$$\Delta t_i = t_i - t_0 \tag{4}$$

where t_0 is some reference epoch. This reference epoch can be arbitrarily chosen, often as the beginning of the time series or a mid-point of the observing campaigns. The other terms are:

(1) $\epsilon_{i,\mathrm{INS}}$ are the residuals to a fit. We assume that each $\epsilon_{i,\mathrm{INS}}$ value is a Gaussian random variable with a zero mean and a variance of $\sigma_i^2 + \sigma_{\mathrm{INS}}^2$, where σ_i^2 is the reported uncertainty of the ith measurement and is the jitter parameter and σ_{INS}^2 represents the excess white noise not included in σ_i^2.

(2) γ_{INS} is the zero-point velocity of each instrument. Each INS can have a different zero point depending on how the radial velocities are measured and how the wavelengths are calibrated.

(3) $\dot{\gamma}$ is a linear trend parameter caused by a long-term acceleration.

(4) The term $\kappa(\Delta t_i)$ is the superposition of k Keplerian signals evaluated at Δt_i. Each k depends on five parameters: the orbital period P_{p}, the semi-amplitude of the signal K_{p}, the mean anomaly $M_{0,\mathrm{p}}$, which represents the phase of the orbit with respect to the periastron of the orbit at t_0, the orbital eccentricity e_{p} that goes from 0 (circular orbit) to 1 (unbound parabolic orbit) and the argument of periastron ω_{p}, which is the angle on the orbital plane with respect to the plane of the sky at which the star goes through the periastron of its orbit (the planet's periastron is at $\omega_{\mathrm{p}} + 180°$). Detailed definitions of the parameters can be found elsewhere[37].

(5) The moving average term

$$\mathrm{MA}_{i,\mathrm{INS}} = \phi_{\mathrm{INS}} \exp\left\{ \frac{t_{i-1} - t_i}{\tau_{\mathrm{INS}}} \right\} \epsilon_{i-1,\mathrm{INS}} \tag{5}$$

is a simple parameterization of the possible correlated noise that depends on the residual of the previous measurement $\epsilon_{i-1,\mathrm{INS}}$. As for the other parameters related to noise in our model, we assume that the parameters of the MA function depend on the instrument; for example the different wavelength ranges used will cause different properties of the instrumental systematic noise. Keplerian and other physical processes also introduce correlations into the data, therefore some degree of degeneracy between the MA terms and the signals of interest is expected. As a result, including an MA term always produces more conservative statistical significance estimates than a model with uncorrelated random noise only. The MA model is implemented through a coefficient ϕ_{INS} and a timescale τ_{INS}. ϕ_{INS} quantifies the strength of the correlation between the i and $i-1$ measurements. It is bound between -1 and 1 to guarantee that the process is stationary (that is, the contribution of the MA term does not arbitrarily grow over time). Exponential smoothing is used to decrease the strength of the correlation exponentially as the difference $t_i - t_{i-1}$ increases[38].

(6) Linear correlations with activity indices can also be included in the model in the following manner

$$\mathrm{A}_{i,\mathrm{INS}} = \sum_{\xi} C_{\xi,\mathrm{INS}}\ \xi_{i,\mathrm{INS}} \tag{6}$$

where ξ runs over all of the activity indices used to model each INS data set (for example, m_2, m_3, S-index and so on, whose descriptions are provided below). To avoid any confusion with other discussions about correlations, we call these $C_{\xi,\mathrm{INS}}$ activity coefficients. Note that each activity coefficient $C_{\xi,\mathrm{INS}}$ is associated with one activity index (ξ_i) obtained simultaneously with the ith radial velocity measurement (for example, chromospheric emission from the Hα line, the second moment of the mean-line profile, the interpolated photometric flux and so on). When fitting a model to the data, an activity coefficient substantially different from 0 indicates evidence of Doppler variability correlated with the corresponding activity index. Formally speaking, these $C_{\xi,\mathrm{INS}}$ correspond to the coefficient of the first-order Taylor expansion of a physical model for the apparent radial velocities as a function of the activity indices and other physical properties of the star.

A simplified version of the same likelihood model is used when analysing time series of activity indices. That is, when searching for periodicities in series other than Doppler measurements, the model will consist of the γ_{INS} zero points, a linear trend term $\dot{\gamma}\Delta t_i$ and a sum of n sinusoids

$$\hat{\kappa}(t_i, \boldsymbol{\theta}) = \sum_{k}^{n} \left(A_k \sin \frac{2\pi\Delta t_i}{P_k} + B_k \cos \frac{2\pi\Delta t_i}{P_k} \right) \tag{7}$$

where each kth sinusoid has three parameters A_k, B_k and P_k instead of the five Keplerian ones. Except for the period parameters and the jitter terms, this model is linear with all the other parameters, which allows a relatively quick computation of the likelihood-ratio periodograms.

Bayesian prior choices. As in any Bayesian analysis, the prior densities of the model parameters have to be selected in a suitable manner (for example see ref. 39). We used uniform and uninformative distributions for most of the parameters apart from a few, possibly important, exceptions. First, as we used a parameter $l = \ln P$ in the Markov chain Monte Carlo samplings instead of the period P directly, the uniform prior density $\pi(l) = c$ for all $l \in [\ln P_0, \ln P_{\mathrm{max}}]$, where P_0 and P_{max} are some minimum and maximum periods, does not correspond to a uniform prior in P. Instead, this prior corresponds to a period prior such that $\pi(P) \propto P^{-1}$ (ref. 40). We made this choice because the period can be considered a scale parameter for which an uninformative prior is one that is uniform in $\ln P$ (ref. 41). We selected the parameter space of the period such that $P_0 = 1$ d and $P_{\mathrm{max}} = T_{\mathrm{obs}}$, where T_{obs} is the time baseline of the combined data sets.

For the semi-amplitude parameter K, we used a $\pi(K) = c$ for all $K \in [0, K_{\mathrm{max}}]$, where K_{max} was selected as 10 m s^{-1} because the r.m.s. of the Doppler series did not exceed 3 m s^{-1} in any of the sets. Following previous works[40,42], we chose the prior for the orbital eccentricities as $\pi(e) \propto \mathcal{N}(0, \Sigma_e^2)$, where e is bound between zero (circular orbit) and 1. We set this $\Sigma_e^2 = 0.3$ to penalize high eccentricities while keeping the option of high e if the data strongly favours it.

We also used an informative prior for the excess white noise parameter of σ_{INS} for each instrument. Based on analyses of a sample of M dwarfs[15], this stellar jitter is typically very close to a value of 1 m s^{-1}. Thus, we used a prior such that $\pi(\sigma_l) \propto \mathcal{N}(\mu_\sigma, \sigma_\sigma^2)$ the parameters were selected as $\mu_\sigma = \sigma_\sigma = 1$ m s^{-1}. Uniform priors were used in all the activity coefficients $C_\xi \in [-C_{\xi,\mathrm{max}}, C_{\xi,\mathrm{max}}]$. For practical purposes, the time series of all activity indices were mean subtracted and normalized to their r.m.s. This choice allows us to select the bounds of the activity coefficients for the renormalized time series as $\hat{C}_{\xi,\mathrm{max}} = 3$ m s^{-1}, so that adding correlation terms does not dramatically increase the r.m.s. of the Doppler time series over the initially measured r.m.s. of <3 m s^{-1} (same argument as for the prior on K). This renormalization is automatically applied by our codes at initialization.

Search for periodicities and statistically significant signals in a frequentist framework. Periodograms are plots that represent a figure of merit derived from a fit against the period of a newly proposed signal. In the case of unevenly sampled data, a very popular periodogram is the Lomb–Scargle periodogram[43,44] and its variants, such as the Floating-mean periodogram[45] or the F-ratio periodogram[46]. In this work we use likelihood ratio periodograms, which represent the improvement of

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

the likelihood statistic when adding a new sinusoidal signal to the model. Owing to intrinsic nonlinearities in the Keplerian/radial velocity modelling, optimizing the likelihood statistic is more computationally intensive than the classic Lomb–Scargle-like periodograms[45,47]. On the other hand, the likelihood function is a more general and well behaved statistic that, for example, allows for the optimization of the noise parameters (such as jitter and the fit correlated noise models at the signal search level). By well-behaved we mean that it has less intrinsic variance compared with other statistics that do not include parameters for the noise such as the χ^2 statistic. Once the maximum likelihood of a model with one additional planet is found (the highest peak in the periodogram), its FAP can then be easily computed[14,48]. In general, an FAP of 1% is needed to claim hints of variability, and a value below 0.1% is considered necessary to claim a statistically significant detection.

Spectroscopic data sets. *New reduction of the UVES M-dwarf programme data.* Between 2000 and 2008, Proxima was observed in the framework of a precision radial velocity survey of M dwarfs in search for extrasolar planets with UVES installed in the Very Large Telescope unit 2. To attain high-precision radial velocity measurements, UVES was self-calibrated with its iodine gas absorption cell operated at a temperature of 70 °C. Image slicer number 3 was chosen, which redistributes the light from a $1'' \times 1''$ aperture along the chosen $0.3''$-wide slit. In this way, a resolving power of $R = 100{,}000$–$120{,}000$ was attained. At the selected central wavelength of 600 nm, the useful spectral range containing iodine (I_2) absorption lines (500—600 nm) falls entirely on the better-quality detector of the mosaic of two 4 K × 2 K CCDs. More details can be found in the several papers from the UVES survey[9,45,49].

The extracted UVES spectra include 241 observations taken through the iodine cell, three template (no iodine) shots of Proxima and three spectra of the rapidly rotating B star HR 5987 that are also taken through the iodine cell and almost consecutive to the three template shots. The B star has a smooth spectrum devoid of spectral features and it was used to calibrate the three template observations of the target. Ten of the iodine observations of Proxima were eliminated due to low exposure levels. The remaining 231 iodine shots of Proxima were taken on 77 nights, typically 3 consecutive shots per night.

The first steps in the processing of the I_2-calibrated data consists of constructing the high signal-to-noise template spectrum of the star without iodine: (1) a custom model of the UVES instrumental profile is generated on the basis of the observations of the B star by forward modelling the observations using a higher-resolution ($R = 700{,}000$–$1{,}000{,}000$) template spectrum of the I_2 cell obtained with the McMath Fourier Transform Spectrometer (FTS) on Kitt Peak; (2) the three template observations of Proxima are then co-added and filtered for outliers; and (3) on the basis of the instrument profile model and wavelength solution derived from the three B star observations, the template is deconvolved with our standard software[10]. After the creation of the stellar template, the 231 iodine observations of Proxima were then run through our standard precision velocity code[8]. The resulting standard deviation of the 231 unbinned observations is 2.58 m s^{-1}, and the standard deviation of the 77 nightly binned observations is 2.30 m s^{-1}, which already suggests an improvement compared to the 3.11 m s^{-1} reported in the original UVES survey reports[49]. All of the UVES spectra (raw) are publicly available in their reduced form via the ESO's archive at http://archive.eso.org/cms.html. Extracted spectra are not produced for this mode of UVES operation, but they are available upon request from the corresponding author.

HARPS GTO. The initial HARPS-Guaranteed Time Observations programme was led by M. Mayor (ESO ID: 072.C-0488). 19 spectra were obtained between May 2005 and July 2008. The typical integration time ranges between 450 s and 900 s.

HARPS M-dwarfs. Led by X. Bonfils and collaborators, this programme consists of ESO programmes 082.C-0718 and 183.C-0437. It produced 8 and 46 measurements, respectively, with integration times of 900 s in almost all cases[50].

HARPS high-cadence. This programme consisted of two 10-night runs (May 2013, and December 2013, ESO ID: 191.C-0505) and was led and executed by several co-authors of this work. Proxima was observed on two runs:

(1) May 2013: 143 spectra obtained on three consecutive nights between 4 May and 7 May and 25 additional spectra between 7 May and 16 May with exposure times of 900 s.

(2) December 2013: 23 spectra obtained between 30 December and 10 January 2014 also with 900 s exposure times.

For simplicity in the presentation of the data and analyses, all HARPS data obtained before 2016 (HARPS GTO, HARPS M-dwarfs, and HARPS high-cadence) are integrated in the HARPS pre-2016 set. The long-term Doppler variability and sparse sampling makes the detection of the Doppler signal more challenging in such a consolidated set than, for example, separating it into subsets of contiguous nights. The latter strategy, however, necessarily requires more parameters (offsets, jitter terms, correlated noise parameters) and arbitrary choices on the sets to be used, producing strong degeneracies and aliasing ambiguities in the determination of the favoured solution (11.2 d was typically favoured, but alternative periods caused by a non-trivial window function at 13.6 d and 18.3 d were also found to be possible). The data taken in 2016 exclusively correspond to the new campaign specifically designed to address the sampling issues.

HARPS PRD campaign. The PRD campaign was executed between 18 January and 30 March 2016. Interruptions of a few nights were anticipated to allow for technical work and other time-critical observations with HARPS. Of the 60 scheduled epochs, we obtained 56 spectra for 54 nights (two spectra were obtained on two of those nights). Integration times were set to 1,200 s, and observations were always obtained at the very end of each night. All of the HARPS spectra (raw, extracted and calibrated frames) are publicly available in their reduced form via ESO's archive at http://archive.eso.org/cms.html.

Spectroscopic indices. Stellar activity can be traced by features in the stellar spectrum. For example, changes in the line-profile shapes (symmetry and width) have been associated with spurious Doppler shifts[18,51]. Chromospheric emission lines are tracers of spurious Doppler variability in the Sun and they are expected to behave similarly for other stars[52]. We describe here the indices measured and used in our analyses.

Measurements of the mean spectral line profiles. The HARPS Data Reduction Software provides two measurements of the mean-line profile shapes derived from the cross-correlation function (CCF) of the stellar spectrum with a binary mask. These are called the bisector span (or BIS) and full-width-at-half-maximum (or FWHM) of the CCF[50]. For very-late-type stars like Proxima, all of the spectral lines are blended, producing a non-trivial shape of the CCF and thus the interpretation of the usual line-shape measurements is not nearly as reliable as in earlier-type stars. We applied the least-squares deconvolution (LSD) technique[53] to obtain a more accurate estimate of the spectral mean line profile. This profile is generated from the convolution of a kernel, which is a model spectrum of line positions and intensities, with the observed spectrum. A description of our implementation of the procedure, applied specifically to crowded M dwarf spectra is described in ref. 54. The LSD profile can be interpreted as a probability function distribution that can then be characterized by its central moments[55]. We computed the second (m_2) and third (m_3) central moments of each LSD profile for each observation. More details of these indices and how they compare with other standard HARPS cross-correlation measurements can be found in ref. 11. To eliminate the correlation of the profile moments with the slope of the spectral energy distribution[11], we corrected the spectral energy distribution and blaze function to match the same spectral energy distribution of the highest signal-to-noise ratio (or S/N) observation obtained with HARPS. Uncertainties were obtained using an empirical procedure as follows: we derived all the m_2 and m_3 measurements of the high-cadence night of 7 May 2013 and fitted a polynomial to each time series. The standard deviation of the residuals to that fit was then assumed to be the expected uncertainty for a $S/N \approx 20$ (at reference echelle aperture number 60), which was the typical value for that night's observations. All other errors were then obtained by scaling this standard deviation by a factor of $\frac{20}{S/N_{\rm obs}}$ for each observation.

Chromospheric indices. Chromospheric emission lines are tracers of spurious Doppler variability in the Sun and they are expected to behave similarly for other stars[52]. We describe here the indices computed and used in our analyses.

Chromospheric Ca II H+K S-index. We calculated the Ca II H+K fluxes following standard procedures[56,57], both the PRD data and the pre-2016 data were treated the same. Uncertainties were calculated from the quadrature sum of the variance in the data used within each bandpass.

Chromospheric Hα emission. This index was measured in a similar way to the *S*-indices, in that we summed the fluxes in the centre of the lines, calculated to be 6,562.808 Å, this time using square bandpasses of 0.678 Å not triangular shapes, and those were normalized to the summed fluxes of two square continuum band regions surrounding each of the lines in the time series. The continuum square bandpasses were centred at 6,550.870 Å and 6,580.309 Å and had widths of 10.75 Å and 8.75 Å, respectively. Again the uncertainties were calculated from the quadrature sum of the variance of the data within the bandpasses.

Photometric data sets. *ASH2.* The ASH2 telescope is a 40 cm robotic telescope with a CCD camera STL11000 2.7 K × 4 K, and a field-of-view (FOV) of 54 × 82 arcmin. Observations were obtained in two narrow-band filters centred on the Hα and S II lines, respectively (Hα is centred on 656 nm, S II is centred on 672 nm, and both filters have a Gaussian-like transmission with a FWHM of 12 nm). The telescope is at SPACEOBS (San Pedro de Atacama Celestial Explorations Observatory), at 2,450 m above sea level, located in the northern Atacama Desert in Chile. This telescope is managed and supported by the Instituto de Astrofísica de Andalucía (Spain). During the present work, only subframes with 40% of the total FOV were used, resulting in a useful FOV of 21.6 × 32.8 arcmin. Approximately 20 images in each band of 100 s of exposure time were obtained per night. In total, 66 epochs of about 100 min each were obtained during this campaign. The number of images collected per night was increased during the second part of the campaign (to about 40 images in each filter per night).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

All CCD measurements were obtained by the method of synthetic aperture photometry using a 2×2 binning. Each CCD frame was corrected in a standard way for dark and flat-fielding. Different aperture sizes were also tested to choose the best one for our observations. A number of nearby and relatively bright stars within the frames were selected as check stars to choose the best ones to be used as comparison stars. After checking their stability, C2 = HD 126625 and C8 = TYC 9010-3029-1, were selected as main comparison stars.

The basic photometric data were computed as the differences in magnitude of the S II and Hα filters for Var-X and C2-X, with Var = Proxima and X = C2+C8)/2. Typical uncertainties of each individual data point are about 6.0 mmag, for both S II and Hα filters. This usually leads to error bars of about 1.3 mmag in the determination of the mean levels of each epoch, assuming 20 points per night once occasional strong activity episodes (such as flares) are removed for the analysis of periodicities. For the analyses, these magnitudes were transformed to relative flux measurements normalized to the mean flux over the campaign.

LCOGT network. The LCOGT is an organization dedicated to time-domain astronomy[13]. To facilitate this, LCOGT operates a homogeneous network of 1 m and 2 m telescopes on multiple sites around the world. The telescopes are controlled by a single robotic scheduler, which is capable of orchestrating complex responsive observing programmes, using the entire network to provide uninterrupted observations of any astronomical target of interest. Each site hosts between one and three telescopes, which are configured for imaging and spectroscopy. The telescopes are equipped with identical instruments and filters, which allows for network redundancy. This means that observations can be seamlessly shifted to alternate sites at any time if the scientific programme requires it, or in the event of poor weather.

Observations for the PRD campaign were obtained on the 1 m network every 24 h in the B and V bands with the Sinistro (4 K × 4 K Fairchild CCD486) cameras, which have a pixel scale of 0.38 arcsec and a FOV of 27×27 arcminutes. In addition, B and V observations were taken every 12 h with the SBIG (4 K × 4 K Kodak KAF-6303E CCD) cameras, with a pixel scale of 0.46 arcsec and a FOV of 16×16 arcminutes. Exposure times ranged between 15 and 40 s and a total of 488 photometrically useful images were obtained during the campaign.

The photometric measurements were performed using aperture photometry with AstroImageJ[58] and DEFOT[59]. The aperture sizes were optimized during the analysis with the aim of minimizing the measurement noise. Proxima and two non-variable comparison stars were identified in a reference image and used to construct the detrended light curves. As with the ASH2 curves, the LCOGT differential magnitudes were transformed to normalized flux to facilitate interpretation and later analyses (see Fig. 3).

Signals in time series. In this section we present a homogeneous analysis of all of the time series (Doppler, activity and photometric) presented in this Letter. In all of the periodograms, the black curve represents the search for a first signal. If one first signal is identified, then a red curve represents the search for a second signal. In the few cases where a second signal is detected, a blue curve represents the search for a third signal. The period of Proxima b is marked with a green vertical line.

Module of the window function. We first present the so-called window function of the three sets under discussion. The window function is the Fourier transform of the sampling[60]. Its module shows the frequencies (or periods) where a signal with 0 frequency (or infinite period) would have its aliases. As shown in Extended Data Fig. 1, both the UVES and HARPS PRD campaigns have a relatively clear window function between 1 and 360 d, meaning that the peaks in periodograms can be interpreted in a very straightforward way (no aliasing ambiguities). For the UVES case, this happens because the measurements were uniformly spread over several years without severe clustering, producing only strong aliases at frequencies beating caused by the usual daily and yearly sampling (peaks at 360, 1, 0.5 and 0.33 d). The window of the PRD campaign is simpler, which is the result of a shorter time span and the uniform sampling of the campaign. On the other hand, the HARPS pre-2016 window function (Extended Data Fig. 1b) contains numerous peaks between 1 and 360 d. This means that signals (for example, activity) in the range of a few hundred days will inject severe interference in the period domain of interest, and explains why this set is where the Doppler signal at 11.2 d is detected with a lower confidence (see Extended Data Fig. 2).

Radial velocities. Here we present likelihood-ratio periodogram searches for signals in the three Doppler time series separately (PRD, HARPS pre-2016 and UVES). They are analysed in the same way as the activity indices to enable direct visual comparison. They differ from the ones presented in the main Letter in the sense that they do not include MA terms and the signals are modelled as pure sinusoids to mirror the analysis of the other time series as close as possible. The resulting periodograms are shown in Extended Data Fig. 2. A signal at 11.2 d was close to detection using UVES data only. However, let us note that the signal was not clearly detectable using the Doppler measurements as provided by the UVES survey[45], and it only became obvious when new Doppler measurements were re-derived using up-to-date iodine codes (see Methods subsection 'New reduction of the UVES M-dwarf programme data'). The signal is weaker in the HARPS pre-2016 data set, but it still appears as a possible second signal after modelling the longer-term variability with a Keplerian at 200 d. Subsets of the HARPS pre-2106 data taken on consecutive nights (for example, HARPS high-cadence runs) also show strong evidence of the same signal. However splitting the data into subsets adds substantial complexity to the analysis and the results become quite sensitive to subjective choices (how to split the data and how to weight each subset). The combination of UVES with all the HARPS pre-2016 (Fig. 1a) already produced an FAP of 1%, but a dedicated campaign was deemed necessary given the caveats with the sampling and activity related variability. The HARPS PRD campaign unambiguously identifies a signal with the same period of approximately 11.2 d. As discussed earlier, the combination of all the data results in a very high significance (FAP $< 10^{-7}$), which implies that the period, but also the amplitude and phase are consistent in all three sets.

Photometry signals and calculation of the FF′ index. The nightly average of the four photometric series was computed after removing the measurements clearly contaminated by flares (see Fig. 3). This produces 43 LCOGT epochs in the B and V bands (80 nights), and 66 ASH2 epochs in both S II and Hα bands (100 nights covered). The precision of each epoch was estimated using the internal dispersion within a given night. All four photometric series show evidence of a long-period signal that is compatible with a photometric cycle at 83 d (probably the rotation) reported before[3]. See periodograms in Extended Data Fig. 3.

In the presence of spots, it has been proposed that spurious variability should be linearly correlated with the value of the normalized flux of the star F, the derivative of the flux F', and the product of FF' (ref. 61) in what is sometimes called the FF' model. To include the photometry in the analysis of the Doppler data, we used the best model fit of the highest-quality light curve (AHS2 S II, which has the lowest post-fit scatter) to estimate F, F' and FF' at the instant of each PRD observation. The relation of F, F' and FF' to the Doppler variability is investigated later in the Bayesian analysis of the correlations.

Width of the mean spectral line as measured by m_2. The m_2 measurement contains a strong variability that closely mirrors the measurements from the photometric time series (see Fig. 3). As in the photometry, the rotation period and its first harmonic (approximately 40 d) are clearly detected in the PRD campaign (see Extended Data Fig. 4). This apparently good match needs to be verified on other stars as it might become a strong diagnostic for stellar activity in M stars. The analysis of the HARPS pre-2016 data also shows very strongly that m_2 is tracing the photometric rotation period of 83 d. The modelling of this HARPS pre-2016 requires a second sinusoid with $P_2 \approx 85$ d, which is peculiar given how close it is to P_1. We suspect this is caused by photospheric features on the surface changing over time.

Asymmetry of the mean spectral lines as monitored by m_3. The periodogram analysis of m_3 of the PRD run suggests a signal at 24 d, which is close to twice the Doppler signal of the planet candidate (see Extended Data Fig. 5). However, line asymmetries are expected to be directly correlated with Doppler signals, not at twice nor integer multiples of the Doppler period. In addition, the peak has an FAP of 5%, which makes it not significantly different from white noise. When looking at the HARPS pre-2016 data, strong beating is observed at 179 and 360 d, which is probably caused by a poorly sampled signal at that period or longer (possibly a magnetic cycle), or some residual systematic effect (possibly contamination by tellurics). In summary, m_3 does not show evidence of any stable signal in the range of interest.

Signal searches in the S-index. Although Hα[51] and other lines such as the sodium doublet (NaD1 and NaD2)[62] have been shown to be the best tracers for activity on M dwarfs, analysing the time series of the S-index is also useful because of its historical use in the long-term monitoring of main-sequence stars[63]. In Extended Data Fig. 6 we show the likelihood ratio periodograms for the S-indices of the HARPS pre-2016 and PRD time series. As can be seen, no signals were found around the 11 d period of the radial velocity signal, however two peaks were found close the 1% FAP threshold with periods of approximately 170 and 340 d. To further test the reality of these possible signals, we performed a Lomb–Scargle periodogram analysis[44] of the combined PRD and pre-2016 HARPS data. This test resulted in the marginal recovery of both the 170 and 340 d peaks seen in the likelihood periodograms, with no emerging peaks around the proposed 11 d Doppler signal. The Lomb–Scargle tests revealed some weak evidence for a signal at much lower periods, around 7 and 30 d.

Given that there is evidence for substantial peaks close to periods of 1 yr, its first harmonic and the lunar period, we also analysed the window function of the time series to check if there was evidence that these peaks are artefacts from the combination of the window function pattern interfering with a real long-period activity signal in the data. The dominant power in the window function is found to increase at periods greater than 100 d, with a forest of strong peaks found in that domain, in comparison to that of sub-100 d periods, which is very flat, representing the noise floor of the time series. This indicates that there are likely to be strong interference patterns from the sampling in this region, and that the signal in the

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

radial velocity data are also not due to the sampling of the data. A similar study in the context of the HARPS M dwarf programme was also done on Proxima[62]. They compared several indices and finally decided to use the intensity of the chromospheric sodium doublet lines. They did not report any notable period at the time, but we suspect this was due to using fewer measurements and not removing the frequent flaring events from the series, which also requires the compilation of a number of observations to reliably identify the outliers caused by flares.

Signal searches in Hα emission. Our likelihood-ratio periodograms for Hα (Extended Data Fig. 7) only show non-significant peaks in the 30–40 d period range. It is important to note that the analyses described above have been performed on multiple versions of the data set, in the sense that we analysed the full data set without removing measurements affected by flaring, then proceeded to reanalyse the activities by dropping data clearly following the flaring periods that Proxima went through when we observed the star. This allowed us to better understand the impact that flares and outliers have on signal interference in the activity indices. Although the distribution of the peaks in periodograms changes somewhat depending on how stringent the cuts are, no emerging peaks were seen close to an 11 d period. Concerning UVES Hα measurements, our likelihood-ratio periodogram did not detect any statistically significant signal.

Further tests on the signal. It has been shown[64] that at least some of the ultraprecise photometric time series measured by the CoRot and Kepler space missions do not have a necessary property to be represented by a Fourier expansion: the underlying function, from which the observations are a sample, must be analytic. An algorithm introduced in ref. 64 can test this property and was applied to the PRD data. The result is that, contrary to the light curves aforementioned, claims that the underlying function is non-analytic does not hold with the information available. Although the null hypothesis cannot be definitively rejected, at least until more data are gathered, our results are consistent with the hypothesis that a harmonic component is present in the Doppler time series.

Flares and radial velocities. Among the high-cadence data from May 2013 with HARPS, two strong flares are fully recorded. During these events, all of the chromospheric lines become prominent in emission, Hα being the one that best traces the characteristic time dependence of flares observed on other stars and the Sun. The spectrum and impact of flares on the radial velocities will be described elsewhere in detail. Relevant to this study, we show that the typical flares on Proxima do not produce correlated Doppler shifts (Extended Data Fig. 8). This justifies the removal of obvious flaring events when investigating signals and correlations in the activity indices.

Complete model and Bayesian analysis of the activity coefficients. A global analysis including all of the radial velocities and indices was performed to verify that the inclusion of correlations would reduce the model probability below the detection thresholds. Equivalently, the Doppler semi-amplitude would become consistent with zero if the Doppler signal was to be described by a linear correlation term. Extended Data Fig. 9 shows the marginalized distributions of linear correlation coefficients with the Doppler semi-amplitude K. Each subset is treated as a separate instrument and has its own zero point, jitter and MA term (coefficient) and its own activity coefficients. In the final model, the timescales of the MA terms are fixed to around 10 d because they were not constrained within the prior bounds, thus compromising the convergence of the chains. The sets under consideration are:

(1) UVES. 70 radial velocity measurements and corresponding Hα emission measurements.

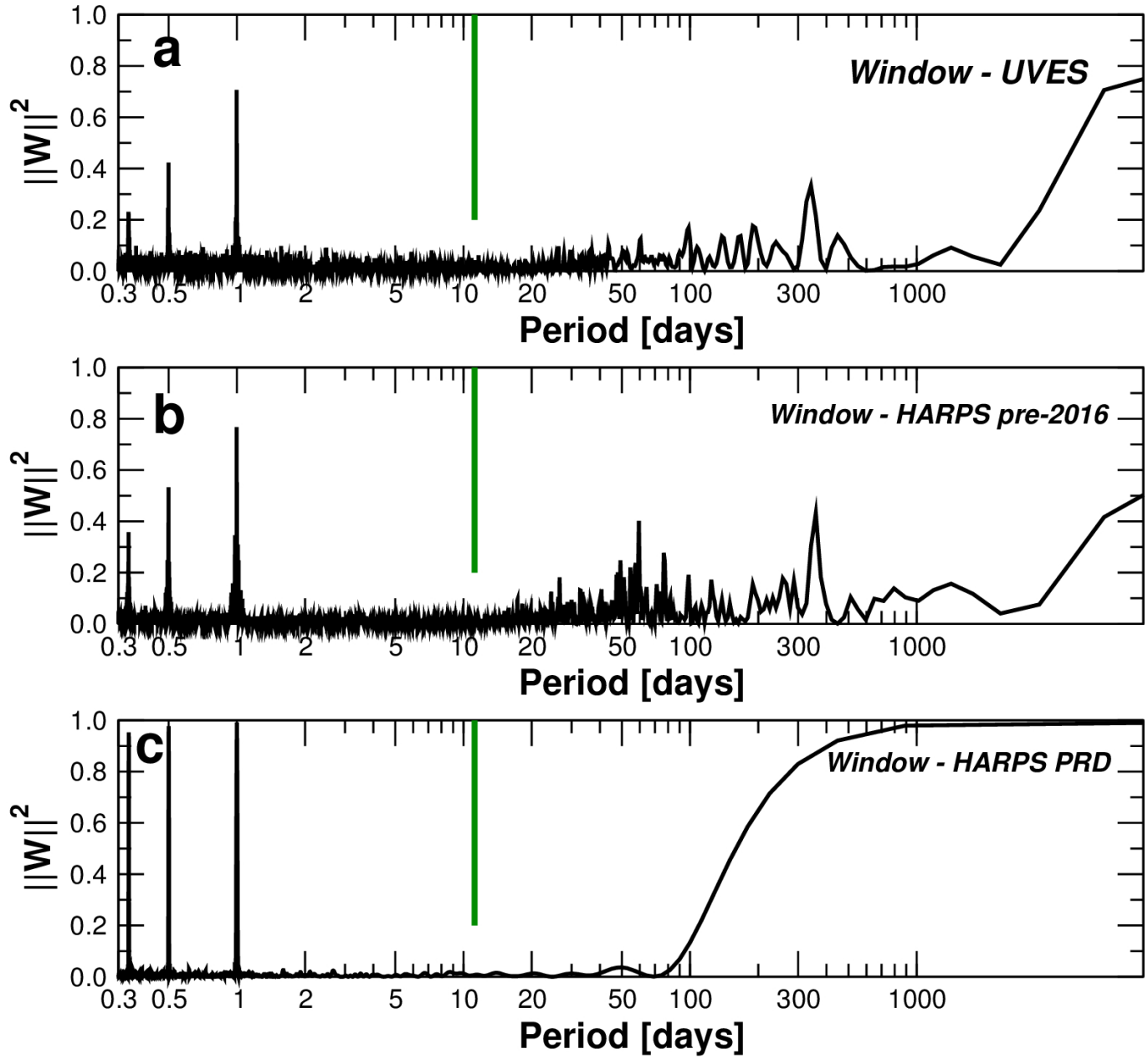
(2) HARPS pre-2016. 90 radial velocity measurements obtained between 2002 and 2014 by several programmes and corresponding spectroscopic indices: m_2, m_3, S-index and the intensities of the Hα and He I lines as measured on each spectrum.

(3) HARPS PRD. 54 Doppler measurements obtained between 18 January–31 March 2016, and the same spectroscopic indices as for the HARPS pre-2016. The values of the F, F' and FF' indices were obtained by evaluating the best fit model to the ASH2 S II photometric series at the HARPS epochs (see Methods subsection 'Photometry signals and calculation of the FF' index').

An activity index is correlated with the radial velocity measurements in a given set if the zero value of its activity coefficient is excluded from the 99% credibility interval. Extended Data Fig. 9 shows the equiprobability contours that contain 50%, 95%, and 99% of the probability density around the mean estimate, and the corresponding 1σ uncertainties in red. Only the F' index (the time derivative of the photometric variability) is substantially different from 0 at high confidence (Extended Data Fig. 9m). Linking this correlation to a physical process requires further investigation. To ensure that such correlations are causally related, one needs a model of the process causing the signal in both the radial velocity and the index, and in the case of the photometry one would need to simultaneously cover more stellar photometric periods to verify that the relation holds over time. Extended Data Table 1 contains a summary of all of the free parameters in the model, including the activity coefficients for each data set.

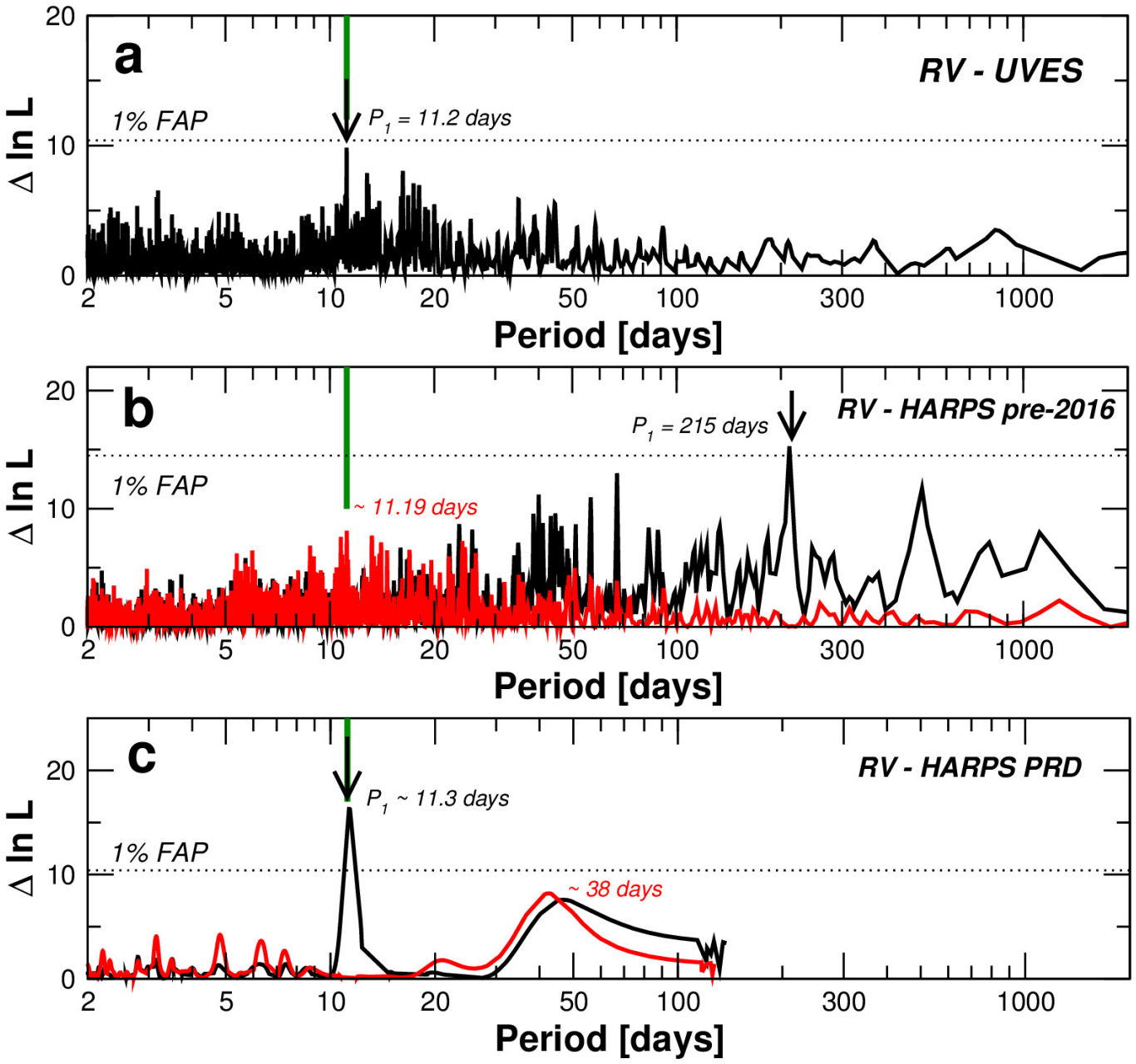
31. Haario, H., Saksman, E. & Tamminen, J. An adaptive Metropolis algorithm. *Bernouilli* **7,** 223 (2001).
32. Tuomi, M. *et al.* Signals embedded in the radial velocity noise: periodic variations in the τ Ceti velocities. *Astron. Astrophys.* **551,** A79 (2013).
33. Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. & Teller, E. Equations of state valculations by fast computing machines. *J. Chem. Phys.* **21,** 1087–1092 (1953).
34. Hastings, W. K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57,** 97–109 (1970).
35. Newton, M. A. & Raftery, A. E. Approximate Bayesian inference with the weighted likelihood bootstrap. *J. R. Stat.Soc. B* **56,** 3–48 (1994).
36. Tuomi, M. A new cold sub-Saturnian candidate planet orbiting GJ 221. *Mon. Not. R. Astron. Soc.* **440,** L1–L5 (2014).
37. Wright, J. T. & Howard, A. W. Efficient fitting of multiplanet Keplerian models to radial velocity and astrometry data. *Astrophys. J. Suppl. Ser.* **182,** 205–215 (2009).
38. Scargle, J. D. Studies in astronomical time series analysis. I: modeling random processes in the time domain. *Astrophys. J. Suppl. Ser.* **45,** 1–71 (1981).
39. Tuomi, M. Evidence for nine planets in the HD 10180 system. *Astron. Astrophys.* **543,** A52 (2012).
40. Tuomi, M. & Anglada-Escudé, G. Up to four planets around the M dwarf GJ 163: sensitivity of Bayesian planet detection criteria to prior choice. *Astron. Astrophys.* **556,** A111 (2013).
41. Berger, J. O. *Statistical Decision Theory and Bayesian Analysis* Section 3.3 (Springer, 1980).
42. Anglada-Escudé, G. *et al.* A dynamically-packed planetary system around GJ 667C with three super-Earths in its habitable zone. *Astron. Astrophys.* **556,** A126 (2013).
43. Lomb, N. R. Least-squares frequency analysis of unequally spaced data. *Astrophys. Space Sci.* **39,** 447–462 (1976).
44. Scargle, J. D. Studies in astronomical time series analysis. II: statistical aspects of spectral analysis of unevenly spaced data. *Astrophys. J.* **263,** 835–853 (1982).
45. Zechmeister, M., Kürster, M. & Endl, M. The M dwarf planet search programme at the ESO VLT + UVES: a search for terrestrial planets in the habitable zone of M dwarfs. *Astron. Astrophys.* **505,** 859–871 (2009).
46. Cumming, A. Detectability of extrasolar planets in radial velocity surveys. *Mon. Not. R. Astron. Soc.* **354,** 1165–1176 (2004).
47. Ferraz-Mello, S. Estimation of periods from unequally spaced observations. *Astron. J.* **86,** 619–624 (1981).
48. Baluev, R. V. Accounting for velocity jitter in planet search surveys. *Mon. Not. R. Astron. Soc.* **393,** 969–978 (2009).
49. Endl, M. & Kürster, M. Toward detection of terrestrial planets in the habitable zone of our closest neighbor: Proxima Centauri. *Astron. Astrophys.* **488,** 1149–1153 (2008).
50. Bonfils, X. *et al.* The HARPS search for southern extra-solar planets. XXXI: the M-dwarf sample. *Astron. Astrophys.* **549,** A109 (2013).
51. Queloz, D. *et al.* No planet for HD 166435. *Astron. Astrophys.* **379,** 279–287 (2001).
52. Robertson, P., Mahadevan, S., Endl, M. & Roy, A. Stellar activity masquerading as planets in the habitable zone of the M dwarf Gliese 581. *Science* **345,** 440–444 (2014).
53. Donati, J.-F. & Brown, S. F. Zeeman–Doppler imaging of active stars. V: sensitivity of maximum entropy magnetic maps to field orientation. *Astron. Astrophys.* **326,** 1135–1142 (1997).
54. Barnes, J. R. *et al.* Red Optical Planet Survey: a new search for habitable Earths in the southern sky. *Mon. Not. R. Astron. Soc.* **424,** 591–604 (2012).
55. Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. *Numerical Recipes in FORTRAN. The Art of Scientific Computing* 2nd edn, Section 4.1 (Cambridge Univ. Press, 1992).
56. Jenkins, J. S. *et al.* An activity catalogue of southern stars. *Mon. Not. R. Astron. Soc.* **372,** 163–173 (2006).
57. Jenkins, J. S. *et al.* Metallicities and activities of southern stars. *Astron. Astrophys.* **485,** 571–584 (2008).
58. Collins, K. A., Kielkopf, J. F. & Stassun, K. G. AstroImageJ: image processing and photometric extraction for ultra-precise astronomical light curves. Preprint at http://arxiv.org/abs/1601.02622 (2016).
59. Southworth, J. *et al.* High-precision photometry by telescope defocussing. VI: WASP-24, WASP-25 and WASP-26. *Mon. Not. R. Astron. Soc.* **444,** 776–789 (2014).
60. Dawson, R. I. & Fabrycky, D. C. Radial velocity planets de-aliased: a new, short period for super-Earth 55 Cnc e. *Astrophys. J.* **722,** 937–953 (2010).
61. Aigrain, S., Pont, F. & Zucker, S. A simple method to estimate radial velocity variations due to stellar activity using photometry. *Mon. Not. R. Astron. Soc.* **419,** 3147–3158 (2012).
62. Gomes da Silva, J. *et al.* Long-term magnetic activity of a sample of M-dwarf stars from the HARPS program. II: activity and radial velocity. *Astron. Astrophys.* **541,** A9 (2012).
63. Baliunas, S. L. *et al.* Chromospheric variations in main-sequence stars. *Astrophys. J.* **438,** 269–287 (1995).
64. Pascual-Granado, J., Garrido, R. & Suárez, J. C. Limits in the application of harmonic analysis to pulsating stars. *Astron. Astrophys.* **581,** A89 (2015).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



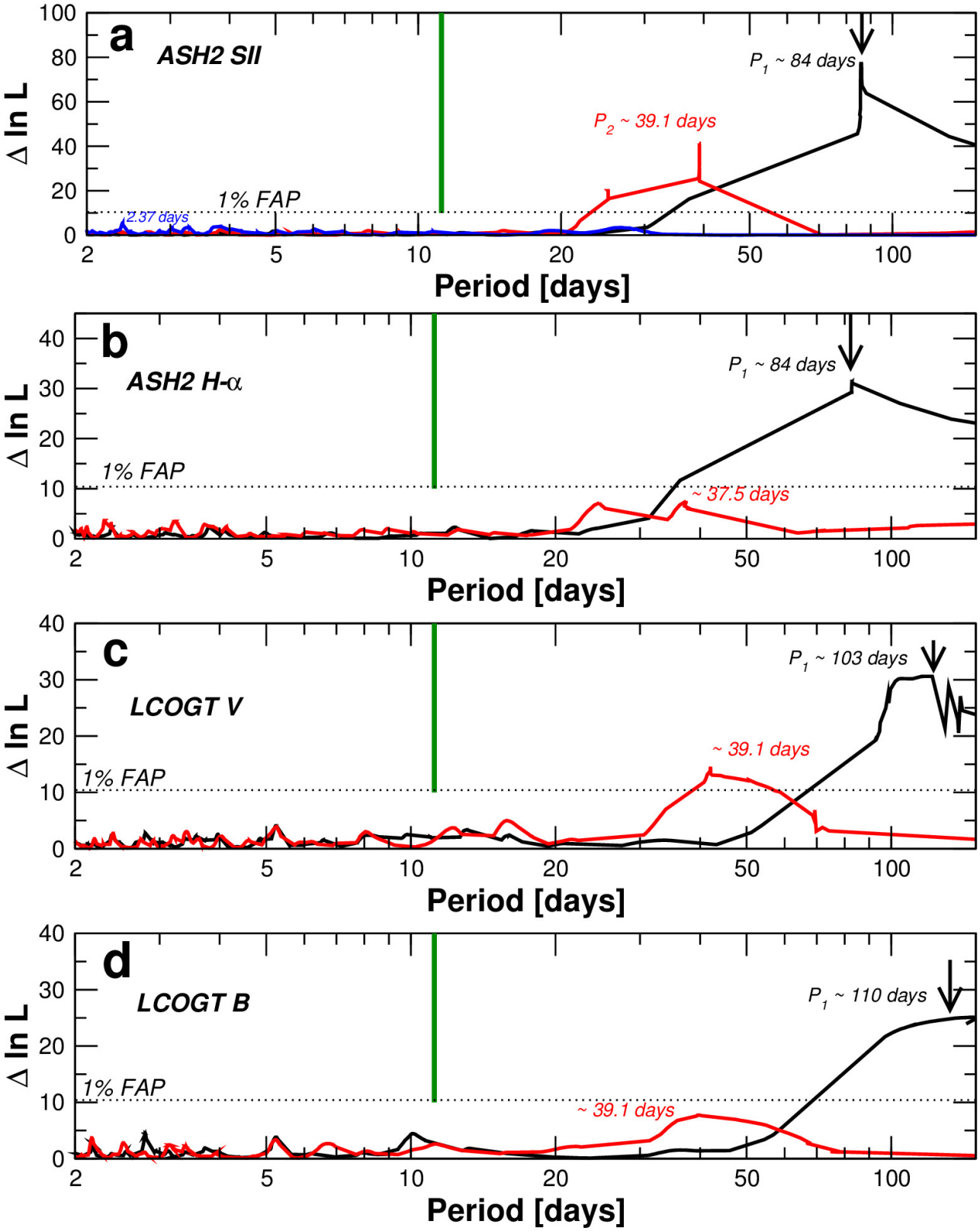
Extended Data Figure 1 | Window function. a–c, Window function of the UVES (**a**), HARPS pre-2016 (**b**) and HARPS PRD (**c**) data sets. The same window function applies to the time series of Doppler and activity data. Peaks in the window function are periods at which aliases of infinite period signals would be expected. The green vertical lines mark the period of the planet candidate at 11.2 d.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



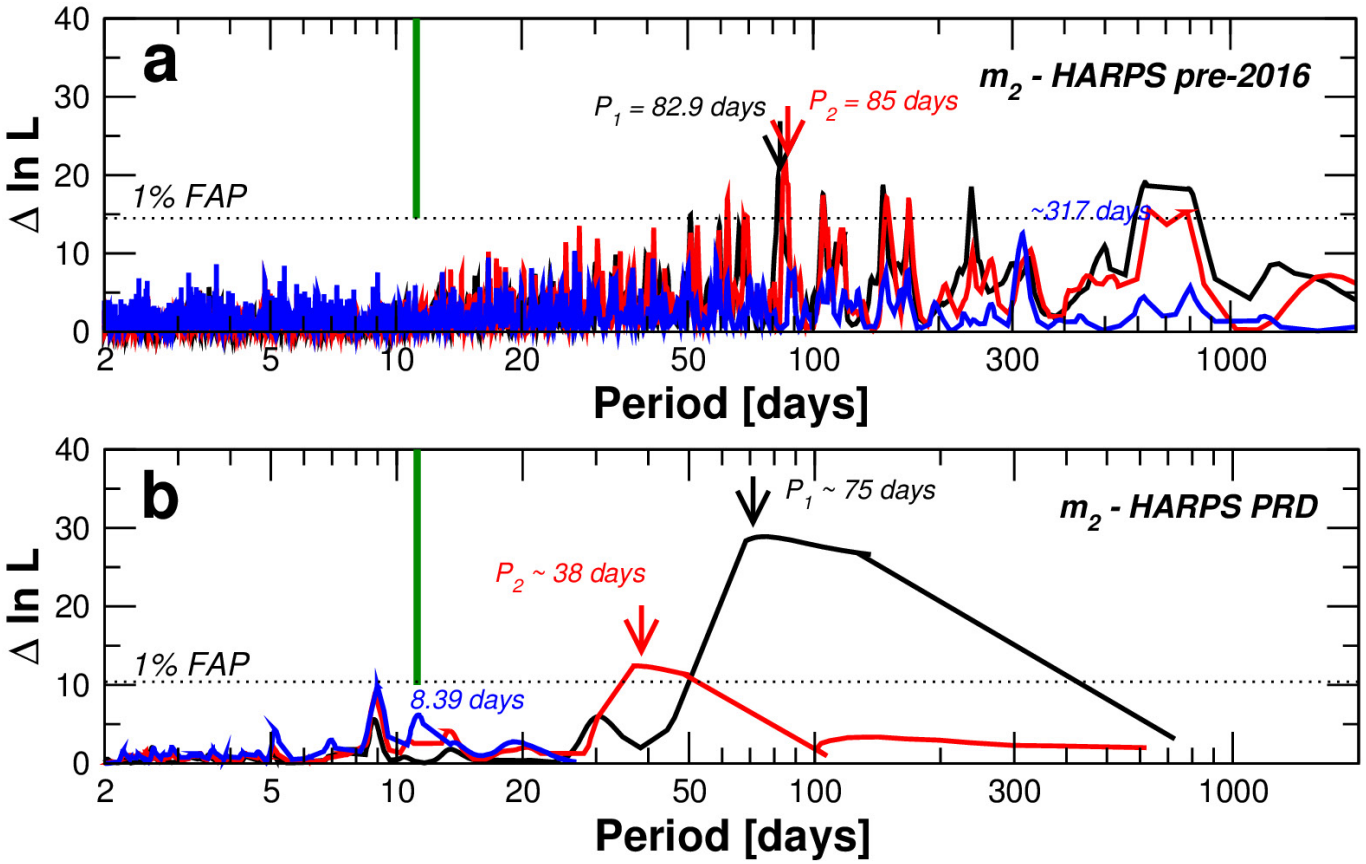
Extended Data Figure 2 | Signal searches on independent radial velocity data sets. a–c, Likelihood-ratio periodograms searches on the radial velocity (RV) measurements of the UVES (**a**), HARPS pre-2016 (**b**) and HARPS PRD (**c**) subsets. The periodogram with all three sets combined is shown in Fig. 1. The black and red lines represent the searches for the first and second signals, respectively. The green vertical lines mark the period of the planet candidate at 11.2 d.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



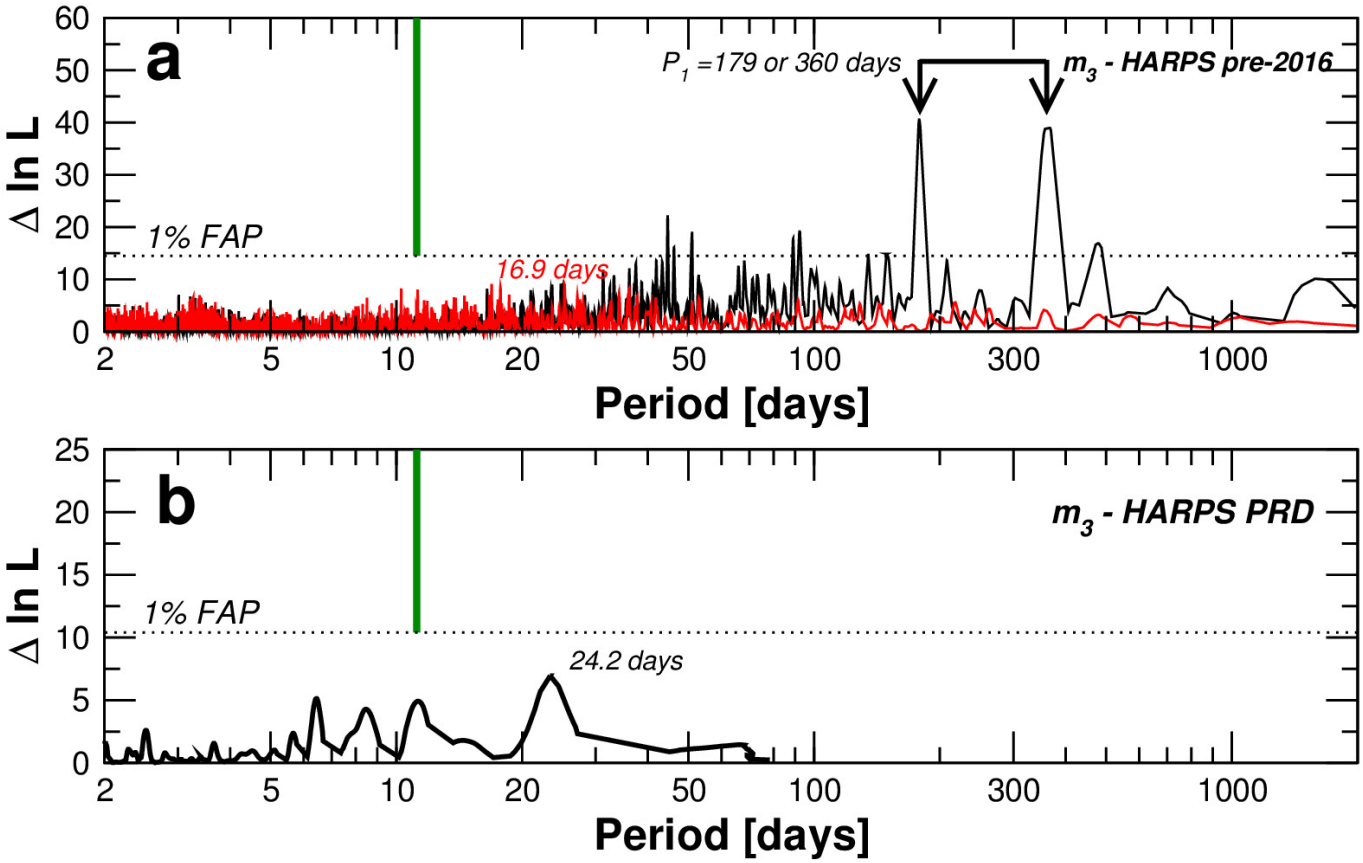
Extended Data Figure 3 | Signal searches on the photometry. **a**–**d**, Likelihood-ratio periodograms searches for signals in each photometric ASH2 photometric band (**a**, **b**) and LCOGT bands (**c**, **d**). The two sinusoid fits to the ASH2 S II series ($P_1 = 84$ d, $P_2 = 39.1$ d) are used later to construct the FF' model to test for correlations of the photometry with the radial velocity data. The black, red and blue lines represent the search for the first, second and third signal respectively. The green vertical lines mark the period of the planet candidate at 11.2 d.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



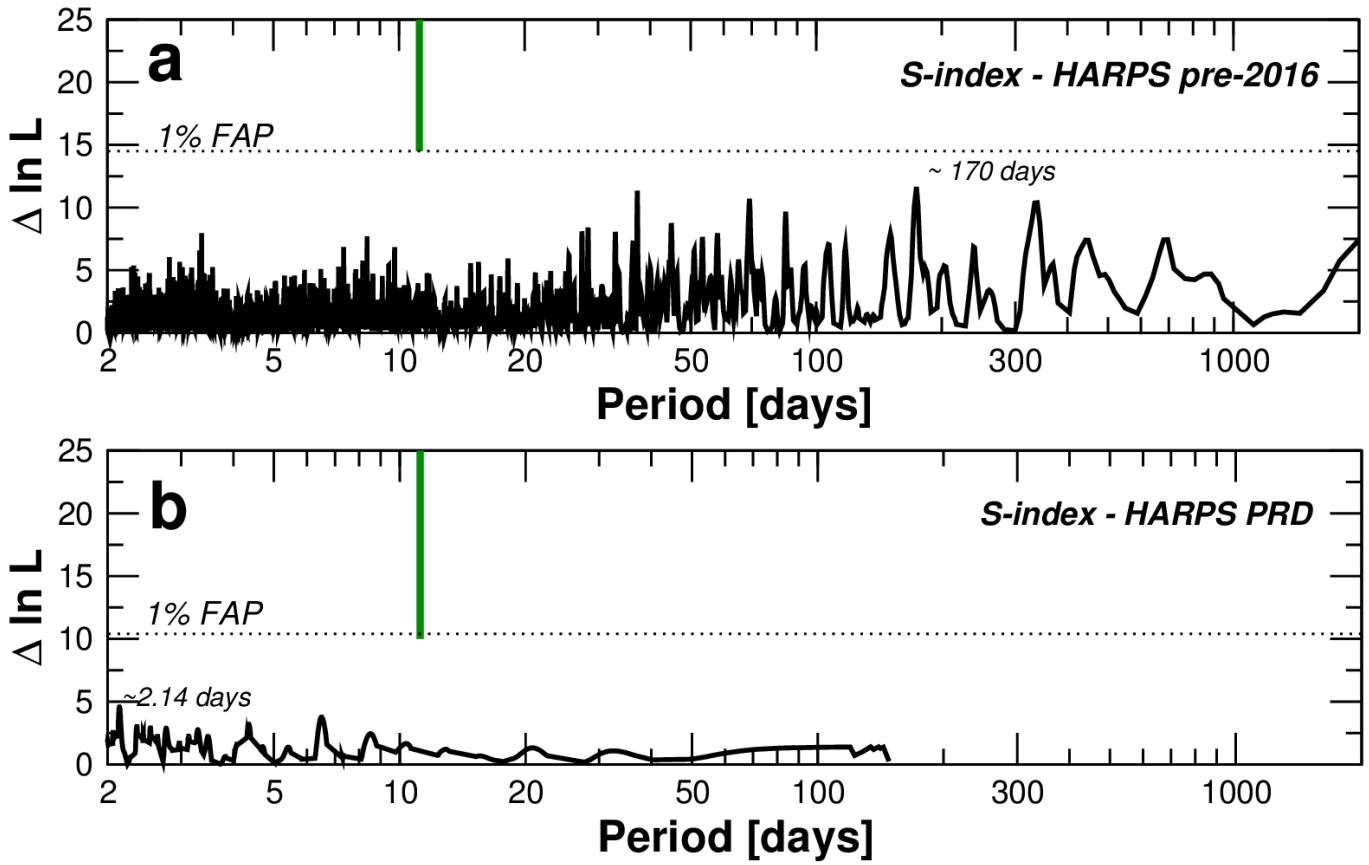
Extended Data Figure 4 | Signal searches on the width of the spectral lines. a, **b**, Likelihood-ratio periodogram searches on the width of the mean spectral line as measured by m_2 for the HARPS pre-2016 (**a**) and HARPS PRD data (**b**). The signals in the HARPS pre-2016 data are comparable to the photometric period reported in the literature and the variability in the HARPS PRD run compares quite well to the photometric variability. The black, red and blue lines represent the search for the first, second and third signal, respectively. The green vertical lines mark the period of the planet candidate at 11.2 d.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



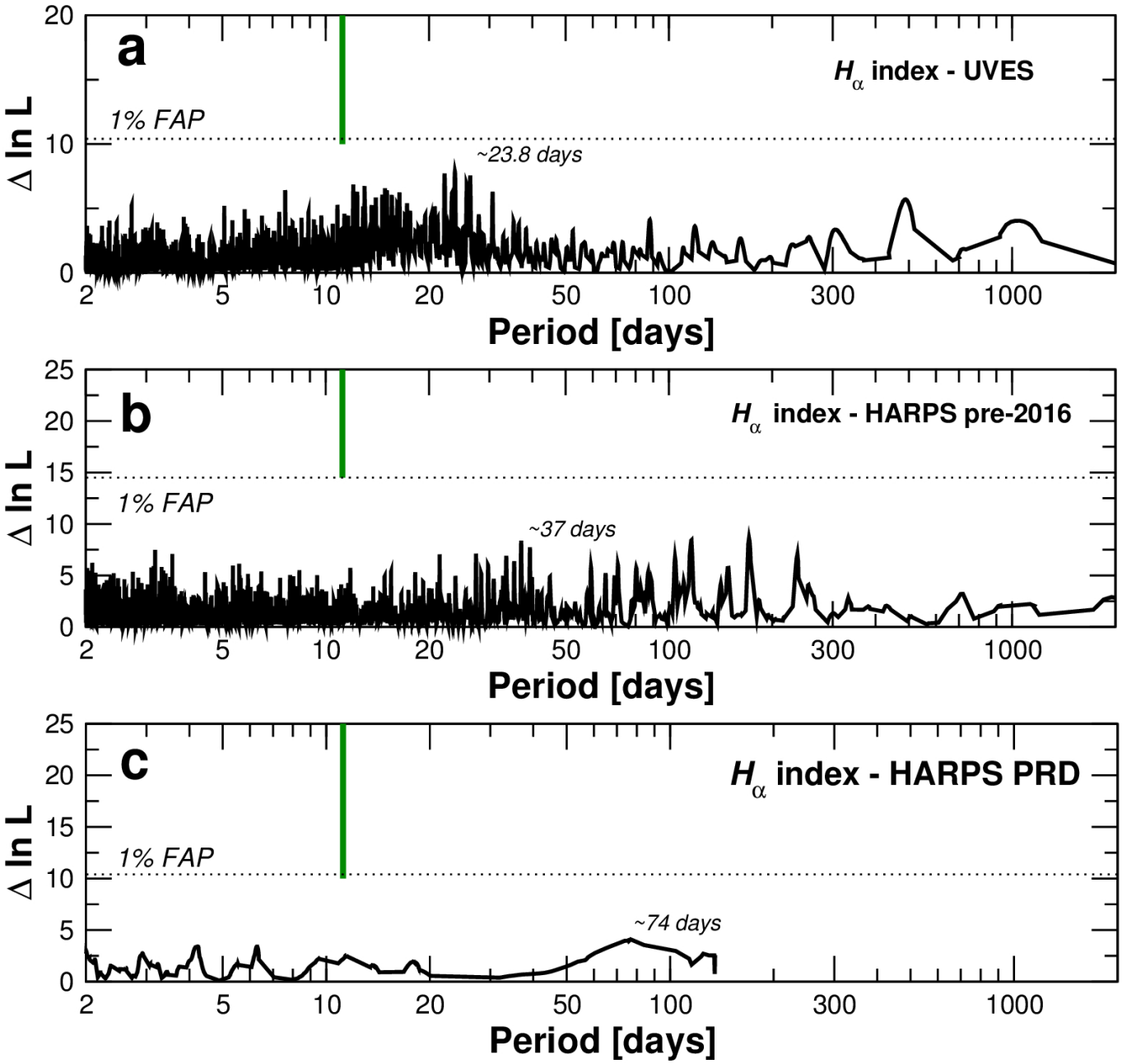
Extended Data Figure 5 | Signal searches on the asymmetry of the spectral lines. a, **b**, Likelihood-ratio periodogram searches on the line asymmetry as measured by m_3 from the HARPS pre-2016 (**a**) and HARPS PRD (**b**) data sets. Signal beating at around 1 yr and 0.5 yr is detected in the HARPS pre-2016 data, which is possibly related to instrumental systematic effects or telluric contamination. No signals are detected above the 1% threshold in the HARPS PRD campaign. The black and red lines represent the search for the first and second signals respectively. The green vertical lines mark the period of the planet candidate at 11.2 d.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



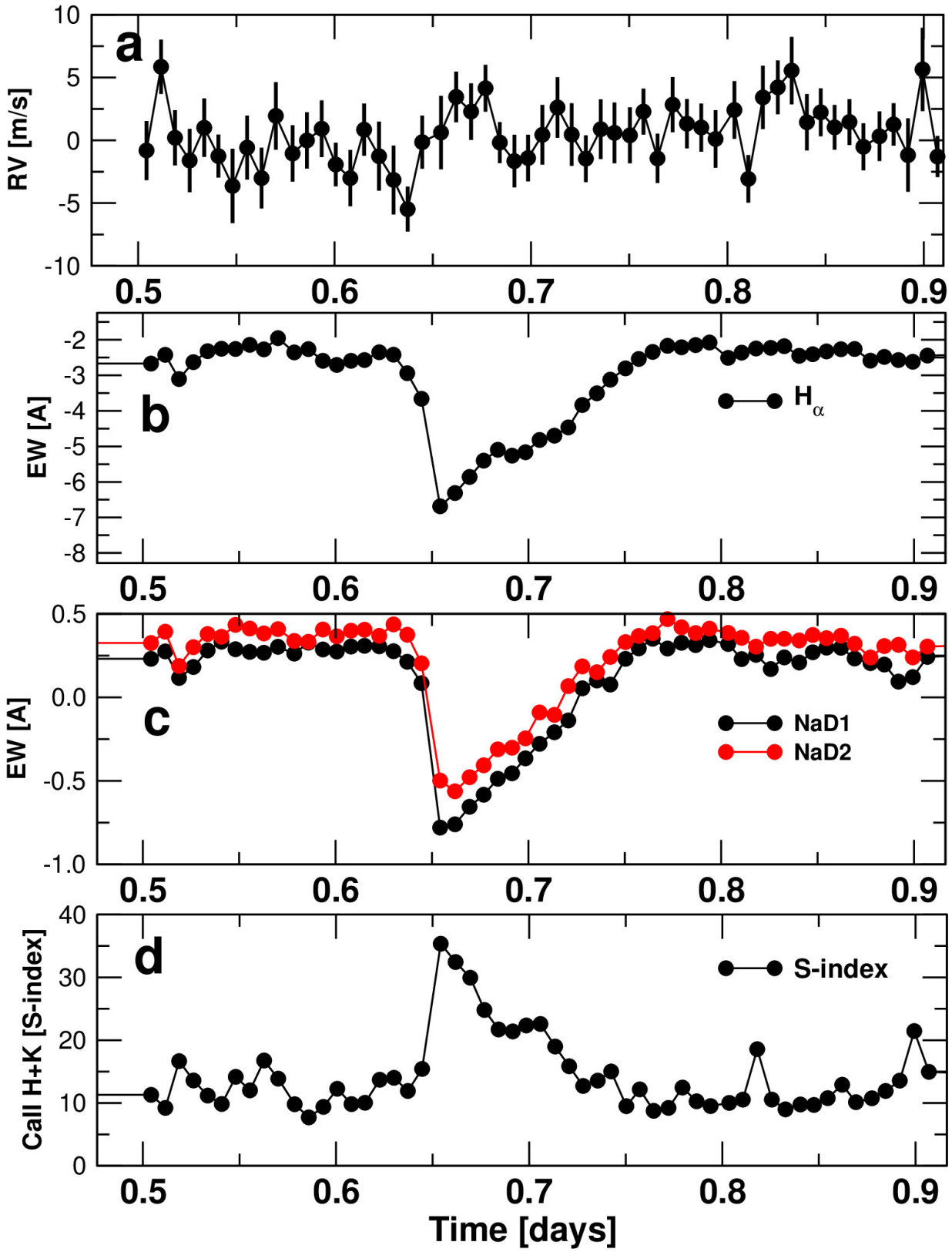
Extended Data Figure 6 | Signal searches on the chromospheric *S*-index. a, **b**, Likelihood-ratio periodogram of the *S*-index from the HARPS pre-2016 (**a**) and HARPS PRD (**b**) campaigns. No signals were detected above the 1% threshold. The green vertical lines mark the period of the planet candidate at 11.2 d.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



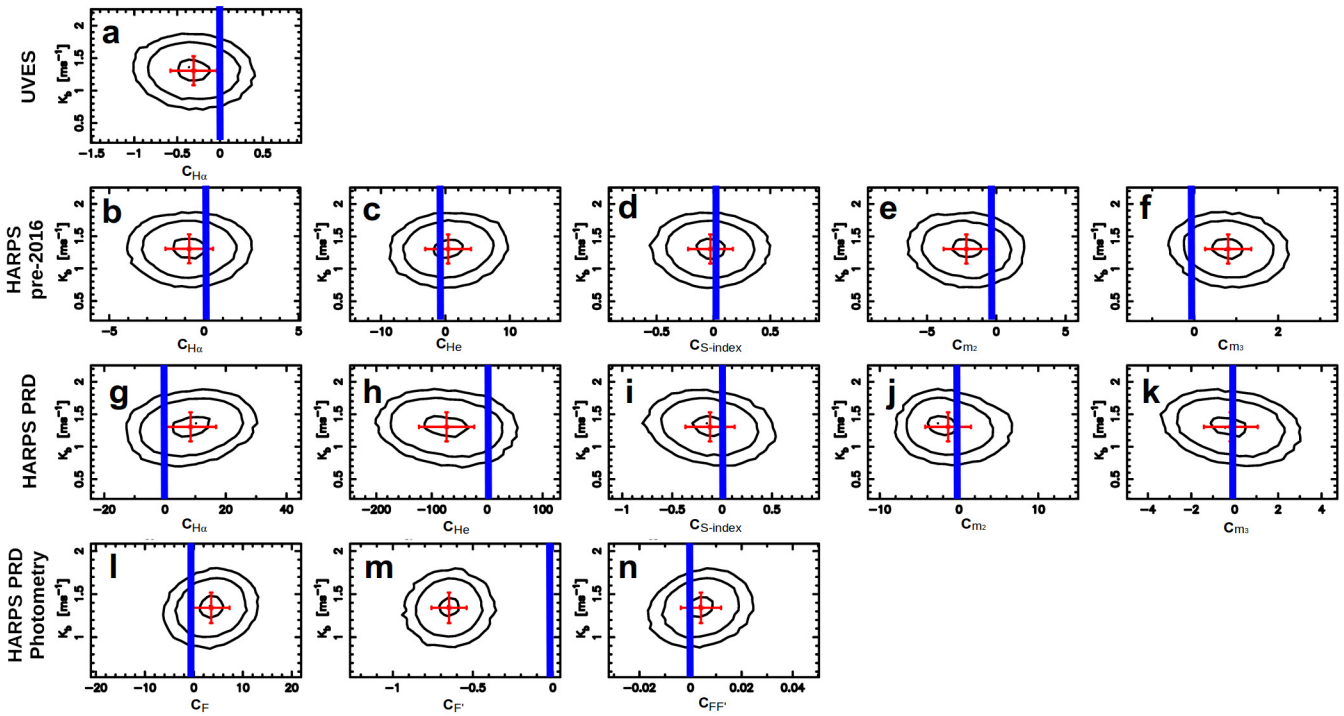
Extended Data Figure 7 | Signal searches on the spectroscopic Hα index. a–c, Likelihood-ratio periodogram searches of Hα intensity from the UVES (**a**), HARPS pre-2016 (**b**) and HARPS PRD (**c**) campaigns. No signals were detected above the 1% threshold. The green vertical lines mark the period of the planet candidate at 11.2 d.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 8 | Radial velocities and chromospheric emission during a flare. a–d, Radial velocities (**a**) and equivalent width measurements of the Hα (**b**), Na doublet lines (**c**) and the *S*-index (**d**) as a function of time during a flare that occurred the night of 5 May 2013. The time axis is days since JD = 245417.0 d. No trace of the flare is observed in the radial velocities. Error bars in the radial velocities correspond to 1σ errors. The formal 1σ errors in the equivalent width measurements are comparable to the size of the points.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 9 | Probability distributions for the activity coefficients versus the signal amplitude. a–n, Marginalized posterior densities of the activity coefficients versus the semi-amplitude of the signal for UVES (**a**), HARPS pre-2016 (**b–f**), HARPS PRD campaign (**g–k**) and the photometric *FF′* indices for the PRD campaign only (**l–n**). Each panel shows equiprobability contours containing 50%, 95% and 99% of the probability density around the mean estimate, and the corresponding standard deviation of the marginalized distribution (1σ) in red. The blue bar shows the zero value of each activity coefficient. Only $C_{F'}$ is found to be substantially different from zero.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 1 | Complete set of model parameters

Parameter	Mean [68% c.i.]	Units
Period	11.186 [11.184, 11.187]	days
Doppler Amplitude	1.38 [1.17, 1.59]	ms^{-1}
Eccentricity	<0.35	–
Mean Longitude	110 [102, 118]	deg
Argument of periastron	310 [-]	deg
Secular acceleration	0.086 [-0.223, 0.395]	$ms^{-1}yr^{-1}$
Noise parameters		
$\sigma_{\rm HARPS}$	1.76 [1.22, 2.36]	ms^{-1}
$\sigma_{\rm PRD}$	1.14 [0.57, 1.84]	ms^{-1}
$\sigma_{\rm UVES}$	1.69 [1.22, 2.33]	ms^{-1}
$\phi_{\rm HARPS}$	0.93 [0.46, 1]	ms^{-1}
$\phi_{\rm PRD}$	0.51 [-0.63, 1]	ms^{-1}
$\phi_{\rm UVES}$	0.87 [-0.02, 1]	ms^{-1}
Activity coefficients*		
UVES		
$C_{\rm H\alpha}$	-0.24 [-1.02, 0.54]	
HARPS pre-2016		
$C_{\rm H\alpha}$	-0.63 [-4.13, 3.25]	
$C_{\rm He}$	1.0 [-9.3, 11.4]	
$C_{\rm S}$	-0.027 [-0.551, 0.558]	
$C_{\rm m_2}$	-1.93 [-6.74, 2.87]	
$C_{\rm m_3}$	0.82 [-0.60, 2.58]	
HARPS PRD		
$C_{\rm H\alpha}$	9.6 [-12.9, 33.3]	
$C_{\rm He}$	-77 [-210, 69]	
$C_{\rm S}$	-0.117 [-0.785, 0.620]	
$C_{\rm m_2}$	-2.21 [-8.86, 7.96]	
$C_{\rm m_3}$	-0.02 [-3.67, 3.44]	
PRD photometry		
$C_{\rm F}$	0.0050 [-0.0183, 0.0284]	
$C_{\rm F'}$	-0.633 [-0.962, -0.304]	
$C_{\rm FF'}$	4.3 [-6.8, 14.8]	

The definition of all of the parameters is given in Methods subsection 'Statistical frameworks and tools'. The values are the maximum *a posteriori* estimates and the uncertainties are expressed as 68% credibility intervals. The reference epoch for this solution is Julian Date $t_0 = 2{,}451{,}634.73146$ d, which corresponds to the first UVES epoch.
*The units of the activity coefficients are metres per second divided by the units of each activity index.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

LETTER

doi:10.1038/nature18949

Extending the lifetime of a quantum bit with error correction in superconducting circuits

Nissim Ofek[1]*, Andrei Petrenko[1]*, Reinier Heeres[1], Philip Reinhold[1], Zaki Leghtas[1]†, Brian Vlastakis[1], Yehan Liu[1], Luigi Frunzio[1], S. M. Girvin[1], L. Jiang[1], Mazyar Mirrahimi[1,2], M. H. Devoret[1] & R. J. Schoelkopf[1]

Quantum error correction (QEC) can overcome the errors experienced by qubits[1] and is therefore an essential component of a future quantum computer. To implement QEC, a qubit is redundantly encoded in a higher-dimensional space using quantum states with carefully tailored symmetry properties. Projective measurements of these parity-type observables provide error syndrome information, with which errors can be corrected via simple operations[2]. The 'break-even' point of QEC—at which the lifetime of a qubit exceeds the lifetime of the constituents of the system—has so far remained out of reach[3]. Although previous works have demonstrated elements of QEC[4–16], they primarily illustrate the signatures or scaling properties of QEC codes rather than test the capacity of the system to preserve a qubit over time. Here we demonstrate a QEC system that reaches the break-even point by suppressing the natural errors due to energy loss for a qubit logically encoded in superpositions of Schrödinger-cat states[17] of a superconducting resonator[18–21]. We implement a full QEC protocol by using real-time feedback to encode, monitor naturally occurring errors, decode and correct. As measured by full process tomography, without any post-selection, the corrected qubit lifetime is 320 microseconds, which is longer than the lifetime of any of the parts of the system: 20 times longer than the lifetime of the transmon, about 2.2 times longer than the lifetime of an uncorrected logical encoding and about 1.1 longer than the lifetime of the best physical qubit (the $|0\rangle_f$ and $|1\rangle_f$ Fock states of the resonator). Our results illustrate the benefit of using hardware-efficient qubit encodings rather than traditional QEC schemes. Furthermore, they advance the field of experimental error correction from confirming basic concepts to exploring the metrics that drive system performance and the challenges in realizing a fault-tolerant system.

Implementing QEC in the laboratory is challenging, requiring a complex system with many qubits. Even for a perfectly realized QEC system of finite size, there will always be unrecoverable errors or failure modes, resulting in an exponential decay of the information over time. In fact, error correction first introduces a hardware overhead penalty, because an uncorrected logical qubit consisting of n physical qubits (for typical first-order codes $n \approx 5$–10; ref. 22) will experience decoherence that is of order n times faster. A central goal of QEC is to suppress the naturally occurring errors and surpass the break-even point, at which the lifetime gain due to error correction is larger than this overhead penalty. These considerations motivate exploring a hardware-efficient approach to QEC, with which it may be more tractable to not only overcome the entire overhead, but to pinpoint the leading limitations to fault-tolerance.

The encoding of logical states as superpositions of Schrödinger-cat states (hereafter, 'cat code') that we implement here is a hardware-efficient scheme that requires fewer physical resources and introduces fewer error mechanisms than do traditional QEC proposals. Designed to operate within a continuous-variable framework[23], the cat code exploits the fact that a coherent state $|\alpha\rangle$ is an eigenstate of the resonator lowering operator $\hat{a}$: $\hat{a}|\alpha\rangle = \alpha|\alpha\rangle$. Using a logical basis comprised of superpositions of cat states, which are eigenstates of photon-number parity, the cat code requires just a single ancilla to monitor the dominant error due to single photon loss induced by resonator energy damping. This error channel gives rise to two effects: deterministic energy decay of the resonator field to vacuum and the stochastic application of $\hat{a}$, which results in a change of photon-number parity of any state within the cat code. The former becomes a limiting factor only at small resonator field amplitudes when coherent state overlap can no longer be neglected and can be addressed through either dissipative pumping approaches[24] or unitary gates. The latter, photon loss, is accompanied by phase shifts of $\pi/2$ about the Z_c axis within the logical space, indicating that by monitoring photon parity as the error syndrome we adhere to the prescriptions for error correction by translating single photon loss into a unitary operation on the encoded state[18,19] (Fig. 1):

$$\hat{a}\left(c_0\left|C_\alpha^+\right\rangle + c_1\left|C_{i\alpha}^+\right\rangle\right) \propto \frac{c_0}{\sqrt{2}}\left(\left|\alpha\right\rangle - \left|-\alpha\right\rangle\right) + i\frac{c_1}{\sqrt{2}}\left(\left|i\alpha\right\rangle - \left|-i\alpha\right\rangle\right)$$
$$= c_0\left|C_\alpha^-\right\rangle + ic_1\left|C_{i\alpha}^-\right\rangle$$

where c_0 and c_1 are arbitrary coefficients satisfying $|c_0|^2 + |c_1|^2 = 1$ and $|C_{(i)\alpha}^{\pm}\rangle = (|(i)\alpha\rangle \pm |-(i)\alpha\rangle)/\sqrt{2}$ (the normalization factor $\sqrt{2}$ holds in the limit of large α (refs 17 and 18)). By detecting photon jumps in real-time with quantum non-demolition parity measurements[21], we learn how the phase relationship between the basis states changes, thereby protecting the encoded qubit from the dominant error channel of the system. The rate of photon jumps scales linearly with the average photon number $\bar{n}$ (ref. 17), which exactly mirrors the aforementioned overhead faced by traditional QEC codes[22,25]. Thus, when implementing the cat code, a central figure of merit when assessing the performance of the QEC system will be the degree to which we can overcome the encoding overhead with the application of fast, repeated parity measurements in time.

We use a 3D circuit quantum electrodynamics (QED) architecture[26] consisting of a single transmon qubit coupled to two waveguide resonators[21,27]. The transmon is used as an ancilla both to provide the error syndrome and to encode and decode the logical states (Supplementary Information, section 3). One resonator stores the logical states while the other is used for ancilla readout and control. The dominant storage–ancilla interaction terms are described by the following Hamiltonian:

$$\hat{H}/\hbar = \omega_s\hat{a}^\dagger\hat{a} + \left(\omega_a - \chi_{sa}\hat{a}^\dagger\hat{a}\right)\left|e\right\rangle\left\langle e\right| - \frac{K}{2}\hat{a}^{\dagger 2}\hat{a}^2$$

with $|e\rangle\langle e|$ the ancilla excited state projector, ω_s and ω_a the storage resonator (henceforth the resonator) and ancilla transition frequencies,

[1]Departments of Physics and Applied Physics, Yale University, New Haven, Connecticut 06510, USA. [2]QUANTIC team, INRIA de Paris, 2 Rue Simone Iff, 75012 Paris, France. †Present address: Centre Automatique et Systèmes, Mines-ParisTech, PSL Research University, 60, boulevard Saint-Michel, 75006 Paris, France.
*These authors contributed equally to this work.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

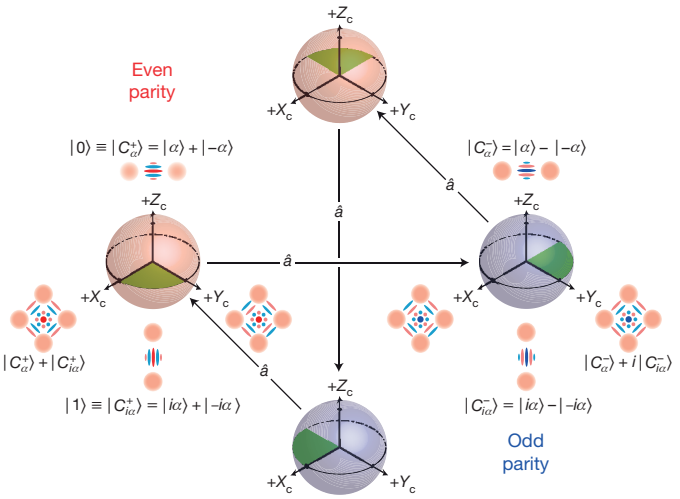


Figure 1 | The cat-code cycle. In the logical encoding of $|0\rangle \equiv |C_\alpha^+\rangle = |\alpha\rangle + |-\alpha\rangle$ and $|1\rangle \equiv |C_{i\alpha}^+\rangle = |i\alpha\rangle + |-i\alpha\rangle$ (normalizations omitted), the two '2-cats' $|C_\alpha^+\rangle$ and $|C_{i\alpha}^+\rangle$ are eigenstates of either even (+) or odd (−) photon-number parity (an '*n*-cat' is a superposition of *n* coherent states). For large enough $|\alpha|$ they are effectively orthogonal to one another. In this basis, the states along the logical axes $+X_c$ and $+Y_c$ are both '4-cats' of even or odd parity as well. The different patterns in the fringes of their cartoon Wigner functions signify the different phase relationship between the basis states. These features allow one to store a qubit in a superposition of 2-cats, $|\psi\rangle = c_0|C_\alpha^+\rangle + c_1|C_{i\alpha}^+\rangle$, and at the same time monitor the parity as the error syndrome without learning anything about c_0 or c_1, the arbitrary coefficients satisfying $|c_0|^2 + |c_1|^2 = 1$. In this example, we choose to encode $|0\rangle$ and $|1\rangle$ in the even-parity basis, although the odd basis can equally be chosen. The loss of a single photon changes not just the parity of the basis states (red shading, even; blue shading, odd), but the phase relationship between them by a factor of *i* as well: $\hat{a}(c_0|C_\alpha^+\rangle + c_1|C_{i\alpha}^+\rangle) = c_0|C_\alpha^-\rangle + ic_1|C_{i\alpha}^-\rangle$. Therefore, after one photon jump, one finds the initial qubit rotated by $\pi/2$ about the logical Z_c axis. With each subsequent application of $\hat{a}$, the encoded state cycles between the even- and odd-parity subspaces, while, owing to each consequent multiplication of the coefficient c_1 by *i*, the encoded information rotates about the Z_c axis by $\pi/2$, as indicated by the rotation of the green shaded slice. Between the stochastic applications of $\hat{a}$, the cat states deterministically decay towards vacuum: $\alpha \rightarrow \alpha e^{-\kappa t/2}$ (not depicted here), indicating that the logical basis changes in time.

respectively, $\chi_{sa}/(2\pi) \approx 1.95$ MHz the dispersive frequency shift, $K/(2\pi) \approx 4.5$ kHz the resonator anharmonicity, or Kerr, and $\hbar$ is the reduced Planck constant. The ancilla has coherence times $T_1 \approx 35\,\mu$s and $T_2 \approx 13\,\mu$s; the resonator has a single-photon Fock state relaxation time $\tau_c \approx 250\,\mu$s and coherence time $T_2^c \approx 330\,\mu$s. To perform high-fidelity single-shot measurements of the ancilla[28], we set the readout resonator to have a 1-MHz bandwidth and use a nearly quantum-limited phase-preserving amplifier, the Josephson parametric converter (JPC)[29], as the first stage of amplification, which allows for a readout fidelity of 99.3%. The error syndrome is measured using a Ramsey-style pulse sequence consisting of two $\pi/2$ pulses applied on the ancilla and separated in time by $\pi/\chi_{sa} \approx 250$ ns (ref. 30). The subsequent ancilla projective readout takes about 700 ns, which includes integration times, cable latencies and feedback delays. A change in the ancilla state after the Ramsey mapping indicates a change in parity, or the loss of one photon (assuming negligible photon excitation) with 98.5% fidelity (Supplementary Information, section 1). The total duration of each error syndrome measurement is just 1 μs, or approximately 0.8% of the average time between photon jumps for cat states with $\bar{n}=2$.

We use a new real-time controller designed to execute programs for quantum control (Supplementary Information, section 5). Every repetition of the program (Fig. 2a–d) begins with the controller encoding one of the six cardinal points on a Bloch sphere in the even logical basis states, enough to perform process tomography of the full QEC system[2]. The number of syndrome measurements and the waiting time t_w between them is set to an optimal value to balance the risk of missing photon jumps and the possibility of measurement back-action on the resonator state due to ancilla T_1 (Supplementary Information, section 4; more details below); for cat states of $\bar{n} \approx 3$, $t_w \approx 13\,\mu$s. The program uses a state machine for adaptive parity monitoring, in which the sign of the second $\pi/2$ pulse in the parity mapping is chosen in real-time to maximize the probability of measuring the ancilla in its ground state $|g\rangle$; this improves measurement reliability by decreasing the probability that the state of the ancilla will change during measurement.

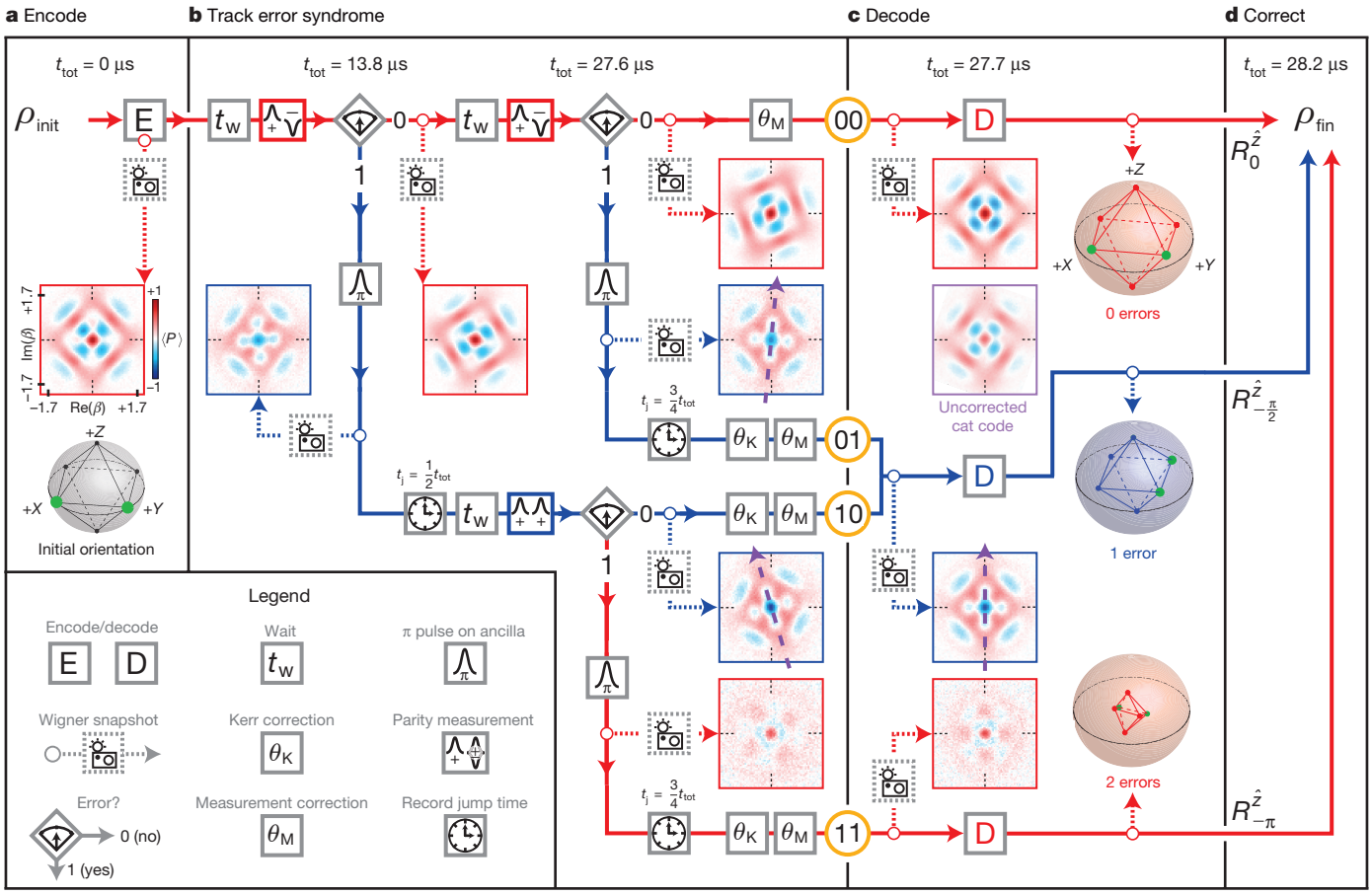
The program stores in memory a record of '0's (no error) and '1's (error) that specifies the monitoring history; Fig. 2b shows the four possibilities for two steps: {00, 01, 10, 11} with probabilities {70.4%, 13.7%, 11.8%, 4.1%}. The asymmetry in the occurrence of 01 and 10 is due to parity measurement infidelity and is well-modelled by a Bayesian analysis (Supplementary Information, section 7). Conditioned on obtaining one of these four records, Wigner tomography provides a striking visual demonstration of the cat code in action. Interference fringes, signatures of quantum coherence[17], continue to be sharp and extremal as the program proceeds in time, as compared with the case of performing no parity monitoring. At each point in the program the tomograms agree well in parity contrast, phase and amplitude as seen in simulations. These levels of predictability highlight the advantages of this hardware-efficient scheme: knowing the Hamiltonian parameters and the measurement fidelity of a single error syndrome is sufficient to encapsulate the evolution of an error-corrected logical qubit.

Two further applications of feedback are necessary to maximize the performance of the QEC system. Owing to the non-commutativity of the Kerr Hamiltonian $\frac{K}{2}\hat{a}^{\dagger 2}\hat{a}^2$ and $\hat{a}$, a photon jump results in a phase shift of the resonator state in phase space proportional to *K* and the jump time t_j: $\theta_K = Kt_j$ (ref. 18). The controller must therefore use the monitoring history, which provides a best-estimate of t_j, to consolidate trajectories of equal error number, yet different error timestamp, into a single effective resonator state in real-time; for example, before decoding, 01 and 10 become a single '1 error' state. The controller also decides in real-time to apply a different set of decoding pulses on the basis of the final parity. Figure 2c shows qubit state tomography of the ancilla after decoding, but before correction, conditioned on the number of errors. The rotation of the six cardinal points by $\pi/2$ for one error and π for two errors indicates that the cat code successfully maps photon-loss errors in the resonator onto a unitary operation on the encoded qubit. Upon completion of execution of the program, the knowledge of how many errors occurred is equivalent to having corrected the state. Although aligning the Bloch spheres of all error trajectories to the same orientation requires a simple phase adjustment on the ancilla drive in the decoding sequence, in this example we instead choose to show the performance of each error case individually. The program thus returns the corrected qubit, now stored again in the ancilla, completing the full QEC cycle.

We benchmark the performance of our QEC system by performing process tomography of the QEC system. We use the chi matrix representation for a single qubit[2], in which state tomography of the output density matrix ρ_{fin} is used to calculate the measured, complex 4×4 chi matrix X^M. The fidelity $F = \mathrm{tr}(X^M X_0)$ is defined as the overlap of X^M with X_0, the chi matrix for the identity operator $\hat{I}$, the ideal case in which a QEC system corrects a state perfectly. Shown in Fig. 2e are the process matrices for the QEC program demonstrated in Fig. 2a–d. The form of X_j^M (X^M for $j=0$, 1 and 2 errors) matches the process matrix for ideal rotations about the *Z* axis (defined by the Pauli matrix σ_z) by $j\pi/2$, $X_{j\pi/2}$. Signatures of developing incoherent mixture are evident from the non-zero values in all diagonal elements.

Moving to an initial encoding size of $\bar{n}_0=2$ to reduce the probability of photon jumps, we implement the cat code with up to six syndrome measurements over approximately 110 μs (Fig. 3a). As in Fig. 2b, each repetition is separated by an optimized waiting time that depends on

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



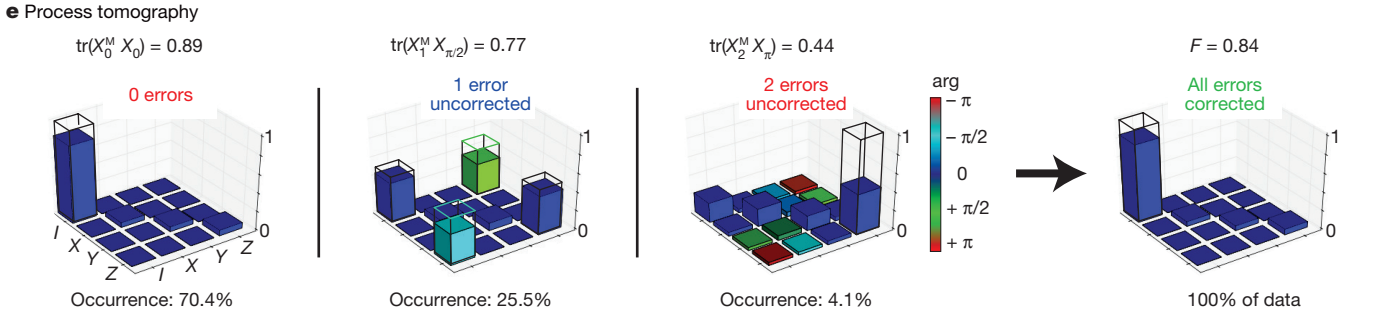


Figure 2 | Example of a two-step quantum trajectory executed by the QEC state machine. a, Six cardinal points on the Bloch sphere ρ_{init} are encoded ('E') from the ancilla onto even-parity resonator states; green markers indicate the initial coordinate-system orientation. A 'Wigner [tomography] snapshot' is shown for $\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) \rightarrow \frac{1}{\sqrt{2}}(|C_\alpha^+\rangle - |C_{i\alpha}^+\rangle)$; $\bar{n}_0$=3; β is the amplitude of varying coherent displacements ($\hat{D}_\beta|0\rangle_f = |\beta\rangle$) and the average parity $\langle P\rangle = \langle \hat{D}_\beta^\dagger \hat{P} \hat{D}_\beta\rangle$, where $\hat{P} = e^{i\pi a^\dagger a}$. **b**, A state machine for adaptive parity monitoring with delays $t_w \approx 13$ µs between each measurement. 'Parity measurement' rectangles show the Ramsey sequence that maps even [+/−] (odd [+/+]) parity onto ancilla $|g\rangle$ ($|e\rangle$); the + or − specifies the sign of the π/2 pulse. Diamonds indicate branching on ancilla measurement (0 → no error, $|g\rangle$; 1 → error, $|e\rangle$); 'π pulse on ancilla' rectangle indicates ancilla reset ($|e\rangle \rightarrow |g\rangle$); clocks indicate recording of the error time t_j. Dashed purple arrows emphasize the phase difference between 10 and 01 due to θ_K. Rotations θ_M are due to cross-Kerr interactions between the readout and storage resonators during ancilla measurements. The parity (Wigner tomogram origin) matches the best estimate (border colour); tomograms match the expected resonator state as seen in simulations. **c**, The feedback aligns all states by changing the phase of subsequent resonator drives (for example, for Wigner tomography or decoding pulses) to account for θ_K and θ_M. Ancilla tomography after decoding shows the expected rotations of π/2 per error about Z (green markers). Different decoding pulses ('D') are chosen in real-time depending on the parity of the final state. **d**, The correction to obtain the final state ρ_{fin} is made via coordinate system rotations ($R^{\hat{z}}$, where $\hat{z}$ is the qubit axis defined by Pauli matrix σ_z) by 0 (for 0 errors), −π/2 (for 1 error) or −π (for 2 errors) in software. **e**, Process tomography results for j = 0, 1 and 2 errors before correction. Ideal $X_{j\pi/2}$ process matrices are shown in wire-outlined bars. Experimental data for X_j^M are shown in solid bars; the values are complex numbers with the amplitude on the vertical axis and an argument specified by the bar colour. Amplitudes of less than 0.01 are not depicted. Process tomography after correction is shown to the right of the arrow. I, the identity matrix; F, fidelity.

the average photon number in the resonator, starting at $t_w \approx 15$ µs and increasing after each subsequent step up to $t_w \approx 25$ µs to account for the decay of the average photon number. Without post-selection, the cat code outperforms the uncorrected transmon with a time constant of exponential decay that is a factor of about 20 higher, indicating that although the coupling between the resonator and transmon is always on, the efficient extraction of error syndromes using an ancilla with inferior coherence properties still allows for substantial gains in lifetime. Moreover, the cat code surpasses the decay of the uncorrected cat code by a factor of about 2.2, demonstrating that applying error correction to the logical encoding makes up for the faster error rates introduced by the hardware overhead. The palpable difference in

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

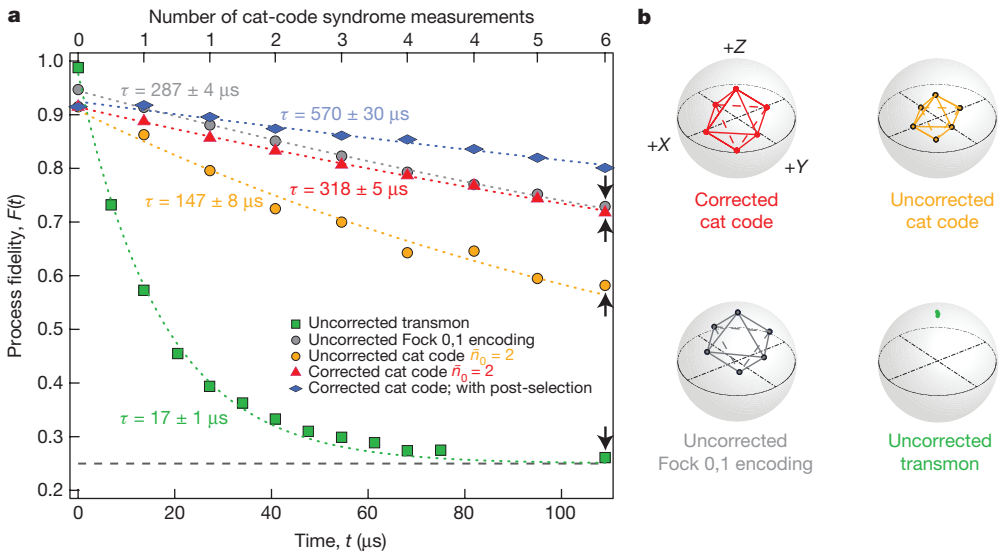


Figure 3 | QEC process tomography. a, To implement QEC, we redundantly encode the qubit in cat states ($\bar{n}_0$=2) and pay the required overhead penalty, which is ubiquitous to QEC. This leads initially to worse performance; the process fidelity ($F(t)$) of the uncorrected cat code (orange circles), where cat states are left to decay freely between encoding and decoding, exhibits faster decay as compared to the Fock states $|0\rangle_f$ and $|1\rangle_f$ (grey circles). Substantial improvements in performance are realized with the full QEC system; the corrected cat code (red triangles) surpasses the uncorrected transmon (green squares) by a factor of about 20, makes up for the QEC overhead by a factor of about 2.2, and offers an improvement over the Fock-state encoding by a factor of about 1.1. With only high-confidence trajectories (blue diamonds), the decay time τ increases to $\tau > 0.5$ ms. The top axis indicates the number of syndrome measurements used for each point in the corrected cat code. Cat code data: 100,000 averages per point; transmon, Fock states: 50,000 averages per point; error bars are smaller than marker sizes. Although no data exhibits strictly single-exponential decay, all curves are well modelled by $F(t) = 0.25 + Ae^{-t/\tau}$ (dotted lines), with τ the decay time of the specific qubit storage scheme and A a fitting constant that is ideally equal to 0.75. $F = 0.25$ (grey dashed line) implies a complete loss of information. Uncertainties are given by the errors (on τ) in the fit. Fluctuations in the uncorrected cat code are explained by the Kerr effect and are reproduced in simulation. **b**, State tomography after approximately 110 μs (corresponding to black arrows in **a**). Transmon and Fock-state Bloch spheres show amplitude damping. Bloch sphere shrinking for the cat code is well-characterized by a depolarization channel. The system substantially benefits from QEC, as seen from the greater definition of each cardinal point in the corrected versus uncorrected case.

contrast between the cardinal points of the uncorrected versus the corrected cat codes after approximately 110 μs (Fig. 3b) demonstrates that with a full QEC system we can enhance the lifetime of a qubit without giving preference to any one direction on the Bloch sphere. By decaying with a time constant that exceeds that of the Fock state by a factor of 1.1, this system reaches the break-even point of QEC.

The history of errors also provides us with a valuable measure of confidence that the result of an error syndrome measurement faithfully reflects the actual error history. Indeed, a 'low-confidence' measurement record that suggests two or more consecutive errors (for example, 11 as in Fig. 2c) has a much lower probability of faithfully reflecting the true error trajectory of the resonator state than does a 'high-confidence' record, wherein a 1 is 'confirmed' by a subsequent 0 (Supplementary Information, section 7). If we accept only high-confidence trajectories, still keeping 80% of the data after 100 μs, then we obtain a decay constant of over half a millisecond. The marked improvement we observe when excluding 'low-confidence' trajectories points to parity measurement infidelity, primarily due to ancilla decoherence, as the dominant limitation on cat-code performance.

An overall analysis of the budget for the lifetime gain for our QEC system is shown in Table 1, which lists the dominant avenues of code failure common to any QEC system and encapsulates the challenges one faces in realizing fault-tolerant QEC. Contributions from the first five entries in Table 1 can be suppressed by measuring more quickly and using a quantum filter to estimate the parity at any given time[21], as seen in the column where $t_w \approx 0$ μs. However, errors due to the ancilla T_1 persist regardless of measurement rate. Owing to its dispersive coupling to the resonator, a change in the energy of the ancilla at an unknown time imparts an unknown rotation to the resonator state in phase space; this is the forward propagation of an error. Measuring the syndrome more frequently only increases the likelihood of ancilla-induced dephasing, necessitating the aforementioned optimized

Table 1 | Failure modes of the corrected logical qubit

Failure mode	Dominant source	Maximum rate, $t_w \approx 0$ μs	Optimal rate, $t_w \approx 20$ μs
		Predicted τ	
Double errors	Cavity $\hat{a} \cdot \hat{a}$	40 ms	1.7 ms
Uncorrectable errors	Cavity $\hat{a}^\dagger$	6 ms	6 ms
Readout error	Transmon T_ϕ	7 ms	2 ms
Ancilla preparation	Transmon $\Gamma_\uparrow$	300 ms	900 μs
Undesired couplings	Cavity $\hat{a}^{\dagger 2}\hat{a}^2$	600 ms	3 ms
Forward propagation	Transmon T_1	200 μs	600 μs
Net lifetime	Predicted	200 μs	320 μs
	Measured	-	318 μs
Gain over uncorrected logical qubit		1.4	2.2
Gain over best physical qubit		0.7	1.1

This table shows the predicted decay time constant (τ) of quantum information stored in a corrected logical qubit using the cat-code paradigm under a scenario in which each individual failure mode is the only source of loss. Dominant modes of failure in the cat code are: double errors ($\hat{a}$ followed by $\hat{a}$) between consecutive syndrome measurements separated by a time t_w; possible errors that the cat code does not address, such as additions of a single photon ($\hat{a}^\dagger$); a failed parity mapping resulting from ancilla dephasing (T_ϕ); incorrect ancilla initialization before syndrome measurement resulting from unknown excitations ($\Gamma_\uparrow$) of the ancilla during t_w; undesired couplings that result in dephasing due to Kerr ($\hat{a}^{\dagger 2}\hat{a}^2$); and ancilla decoherence that directly propagates to unrecoverable errors in the resonator state, which is a result of ancilla decay or excitation (T_1). Two different measurement strategies are shown for an initial $\bar{n}_0$=2: as quickly as possible ($t_w \approx 0$ μs) and the optimal monitoring time ($t_w \approx 20$ μs). The lowest two rows show the multiplicative gains of cat-code performance over the decay constants of the uncorrected logical qubit (147 μs) and the best physical qubit of the system (287 μs, Fock states $|0\rangle_f, |1\rangle_f$). These gains reflect the combined effects of all loss channels acting together. We predict all numbers using an analytical model derived in Supplementary Information, section 6, and show that for the net gains the failure modes do not contribute independently. Using the optimal measurement strategy, we find that the predicted gains in lifetime over the uncorrected logical qubit and over the Fock state encoding match the measured performance of the corrected cat code (318 μs) shown in Fig. 3. Lifetimes of at least 6 ms would be possible if the forward propagation of errors from the syndrome measurements were abated.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

measurement cadence that sets the delay between syndrome measurements from $t_w \approx 0\,\mu s$ to, on average, $t_w \approx 20\,\mu s$ for $\bar{n}_0 = 2$. We therefore see that when designing a QEC system, sources of decoherence beyond double-errors per round of correction can motivate substantially slower measurement rates. However, because the cat code performs at the break-even point even in the presence of all of these sources of loss, we are optimistic about the prospect of realizing a fault-tolerant QEC system. Indeed, supplementing the cat code with a scheme that abates ancilla back-action promises to allow increased error syndrome measurement rates (a lower t_w) and thus greater gains in lifetime.

Our results show that QEC can actually protect an unknown bit of quantum information, and extend its lifetime by active means. Furthermore, we demonstrate the crucial role of real-time feedback with pulses that depend on the evolution of the quantum system, an addition to the experimental setup that greatly improves error correction performance and allows us to realize the cat code at the break-even point of QEC. Future goals include combining the cat code with mechanisms to re-inflate cat state amplitudes[24] and to equip the parity monitoring protocol to handle changes in ancilla energy, thereby addressing issues of non-fault-tolerance. With such capabilities we can then move beyond using the cat code as a quantum memory only and begin coupling multiple resonators together to demonstrate operations between error-corrected logical qubits[19]. Our results motivate the adoption of hardware-efficient QEC schemes and demonstrate the promise of cat states as a basis for future quantum computing applications.

Received 12 February; accepted 8 June 2016.
Published online 20 July 2016.

1. Shor, P. W. Scheme for reducing decoherence in quantum computer memory. *Phys. Rev. A* **52,** R2493–R2496 (1995).
2. Nielsen, M. & Chuang, I. *Quantum Computation and Quantum Information* (Cambridge Univ. Press, 2010).
3. Devoret, M. H. & Schoelkopf, R. J. Superconducting circuits for quantum information: an outlook. *Science* **339,** 1169–1174 (2013).
4. Cory, D. *et al.* Experimental quantum error correction. *Phys. Rev. Lett.* **81,** 2152–2155 (1998).
5. Knill, E., Laflamme, R., Martinez, R. & Negrevergne, C. Benchmarking quantum computers: the five-qubit error correcting code. *Phys. Rev. Lett.* **86,** 5811–5814 (2001).
6. Chiaverini, J. *et al.* Realization of quantum error correction. *Nature* **432,** 602–605 (2004).
7. Schindler, P. *et al.* Experimental repetitive quantum error correction. *Science* **332,** 1059–1061 (2011).
8. Waldherr, G. *et al.* Quantum error correction in a solid-state hybrid spin register. *Nature* **506,** 204–207 (2014).
9. Taminiau, T. H., Cramer, J., van der Sar, T., Dobrovitski, V. V. & Hanson, R. Universal control and error correction in multi-qubit spin registers in diamond. *Nat. Nanotechnol.* **9,** 171–176 (2014).
10. Cramer, J. *et al.* Repeated quantum error correction on a continuously encoded qubit by real-time feedback. *Nat. Commun.* **7,** 11526 (2016).
11. Pittman, T. B., Jacobs, B. C. & Franson, J. D. Demonstration of quantum error correction using linear optics. *Phys. Rev. A* **71,** 052332 (2005).
12. Aoki, T. *et al.* Quantum error correction beyond qubits. *Nat. Phys.* **5,** 541–546 (2009).
13. Reed, M. D. *et al.* Realization of three-qubit quantum error correction with superconducting circuits. *Nature* **482,** 382–385 (2012).
14. Kelly, J. *et al.* State preservation by repetitive error detection in a superconducting quantum circuit. *Nature* **519,** 66–69 (2015).
15. Córcoles, A. D. *et al.* Demonstration of a quantum error detection code using a square lattice of four superconducting qubits. *Nat. Commun.* **6,** 6979 (2015).
16. Ristè, D. *et al.* Detecting bit-flip errors in a logical qubit using stabilizer measurements. *Nat. Commun.* **6,** 6983 (2015).
17. Haroche, S. & Raimond, J.-M. *Exploring the Quantum: Atoms, Cavities, and Photons* (Oxford Univ. Press, 2006).
18. Leghtas, Z. *et al.* Hardware-efficient autonomous quantum memory protection. *Phys. Rev. Lett.* **111,** 120501 (2013).
19. Mirrahimi, M. *et al.* Dynamically protected cat-qubits: a new paradigm for universal quantum computation. *New J. Phys.* **16,** 045014 (2014).
20. Vlastakis, B. *et al.* Deterministically encoding quantum information using 100-photon Schrödinger cat states. *Science* **342,** 607–610 (2013).
21. Sun, L. *et al.* Tracking photon jumps with repeated quantum non-demolition parity measurements. *Nature* **511,** 444–448 (2014).
22. Steane, A. M. Error correcting codes in quantum theory. *Phys. Rev. Lett.* **77,** 793–797 (1996).
23. Braunstein, S. L. & van Loock, P. Quantum information with continuous variables. *Rev. Mod. Phys.* **77,** 513–577 (2005).
24. Leghtas, Z. *et al.* Confining the state of light to a quantum manifold by engineered two-photon loss. *Science* **347,** 853–857 (2015).
25. Fowler, A. G., Mariantoni, M., Martinis, J. M. & Cleland, A. N. Surface codes: towards practical large-scale quantum computation. *Phys. Rev. A* **86,** 032324 (2012).
26. Wallraff, A. *et al.* Strong coupling of a single photon to a superconducting qubit using circuit quantum electrodynamics. *Nature* **431,** 162–167 (2004).
27. Paik, H. *et al.* Observation of high coherence in Josephson junction qubits measured in a three-dimensional circuit QED architecture. *Phys. Rev. Lett.* **107,** 240501 (2011).
28. Vijay, R., Slichter, D. H. & Siddiqi, I. Observation of quantum jumps in a superconducting artificial atom. *Phys. Rev. Lett.* **106,** 110502 (2011).
29. Bergeal, N. *et al.* Phase-preserving amplification near the quantum limit with a Josephson ring modulator. *Nature* **465,** 64–68 (2010).
30. Bertet, P. *et al.* Direct measurement of the Wigner function of a one-photon Fock state in a cavity. *Phys. Rev. Lett.* **89,** 200402 (2002).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank K. Sliwa, A. Narla and L. Sun for discussions. This research was supported by the US Army Research Office (W911NF-14-1-0011). A.P. was supported by the National Science Foundation (NSF) (PHY-1309996). S.M.G. acknowledges additional support from NSF DMR-1301798. Facilities use was supported by the Yale Institute for Nanoscience and Quantum Engineering (YINQE), the Yale SEAS cleanroom and the NSF (MRSECDMR 1119826).

Author Contributions A.P. and N.O. performed the experiment and analysed the data. N.O. designed and built the feedback architecture with help from Y.L. under the supervision of R.J.S. and M.H.D. R.H. and P.R. developed the optimal control pulses. M.M., Z.L., L.J., S.M.G. and B.V. provided theoretical support. R.H. and L.F. fabricated the transmon qubit. R.J.S. supervised the project. A.P., N.O., L.F. and R.J.S. wrote the manuscript with feedback from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.P. (andrei.o.petrenko@gmail.com), N.O. (nissim.of@gmail.com) or R.J.S. (robert.schoelkopf@yale.edu).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

LETTER

doi:10.1038/nature19062

Molecular modifiers reveal a mechanism of pathological crystal growth inhibition

Jihae Chung[1], Ignacio Granja[2], Michael G. Taylor[3], Giannis Mpourmpakis[3], John R. Asplin[2] & Jeffrey D. Rimer[1]

Crystalline materials are crucial to the function of living organisms, in the shells of molluscs[1–3], the matrix of bone[4], the teeth of sea urchins[5], and the exoskeletons of coccoliths[6]. However, pathological biomineralization can be an undesirable crystallization process associated with human diseases[7–9]. The crystal growth of biogenic, natural and synthetic materials may be regulated by the action of modifiers, most commonly inhibitors, which range from small ions and molecules[10,11] to large macromolecules[12]. Inhibitors adsorb on crystal surfaces and impede the addition of solute, thereby reducing the rate of growth[13,14]. Complex inhibitor–crystal interactions in biomineralization are often not well elucidated[15]. Here we show that two molecular inhibitors of calcium oxalate monohydrate crystallization—citrate and hydroxycitrate—exhibit a mechanism that differs from classical theory in that inhibitor adsorption on crystal surfaces induces dissolution of the crystal under specific conditions rather than a reduced rate of crystal growth. This phenomenon occurs even in supersaturated solutions where inhibitor concentration is three orders of magnitude less than that of the solute. The results of bulk crystallization, *in situ* atomic force microscopy, and density functional theory studies are qualitatively consistent with a hypothesis that inhibitor–crystal interactions impart localized strain to the crystal lattice and that oxalate and calcium ions are released into solution to alleviate this strain. Calcium oxalate monohydrate is the principal component of human kidney stones[16–19] and citrate is an often-used therapy[20], but hydroxycitrate is not. For hydroxycitrate to function as a kidney stone treatment, it must be excreted in urine. We report that hydroxycitrate ingested by non-stone-forming humans at an often-recommended dose leads to substantial urinary excretion. *In vitro* assays using human urine reveal that the molecular modifier hydroxycitrate is as effective an inhibitor of nucleation of calcium oxalate monohydrate nucleation as is citrate. Our findings support exploration of the clinical potential of hydroxycitrate as an alternative treatment to citrate for kidney stones.

Figure 1a illustrates the habit of calcium oxalate monohydrate (COM) crystals with indexed faces. Here, we examine COM crystallization in the presence of two inhibitors with nearly identical structure (Fig. 1b): citrate (CA) and hydroxycitrate (HCA). These two molecules differ only by a single alcohol group, yet this subtle difference markedly alters their specificity for binding to COM crystal surfaces. Scanning electron microscopy (SEM) images of COM crystals (Fig. 1c) reveal that CA alters growth in the *c* direction (Fig. 1d), which is consistent with results from prior atomic force microscopy (AFM) measurements[21]. Conversely, HCA binds to the apical tips (that is, the $\{12\bar{1}\}$ and {021} surfaces) and generates diamond-shaped crystals (Fig. 1e). Inhibitor–crystal interactions reduce the *c*/*b* aspect ratio (Fig. 1f) as well as the [100] dimension (Fig. 1g) of COM crystals. Kinetic studies using an ion selective electrode (ISE) to track the temporal depletion of free Ca^{2+} ions[22] in supersaturated calcium oxalate solution (Extended Data Fig. 1) indicate that HCA is a more potent growth inhibitor (Fig. 1h). Note that the third dissociation constant of CA ($pK_a = 6.4$) is close to the pH of COM growth solution in this study (pH 6.2), thus indicating a distribution of CA species with charge of either −3 or −2 (Extended Data Fig. 2a). To ensure that we are comparing the effects of fully dissociated HCA and CA, we performed ISE measurements of COM growth at pH 8.0 (Extended Data Fig. 2b), well above the pK_a of CA, and found that HCA is the more effective inhibitor, irrespective of solution alkalinity (see Methods for a detailed discussion). Increasing concentrations of both CA and HCA leads to a maximum 60% inhibition of COM crystallization. This net effect reflects the reduction in crystal growth rate as well as the potential inhibition of COM nucleation. To assess the latter, we measure the number of crystals collected on substrates per unit area, or the crystal number density ρ_{COM}. As shown in Fig. 1i and Extended Data Fig. 3, HCA and CA both inhibit COM nucleation, resulting in 62% ± 15% and 39% ± 17% reductions in ρ_{COM}, respectively.

Inhibitor interactions with hillocks presented on the surfaces of COM crystals reduce the rate of step advancement. *In situ* AFM measurements confirm that CA exhibits specificity for steps on the basal (100) surface of COM crystals that advance in the *c* direction. Time-resolved images also show that a CA concentration of $>1\,\mu g\,ml^{-1}$ reduces interstep distance, decreases step velocity, and generates protrusions on steps (Fig. 1j). Continuous imaging of COM (100) hillocks reveal that $1\,\mu g\,ml^{-1}$ HCA reduces interstep distances near the origin of screw dislocations (arrow in Fig. 1k) and slows the rate of step advancement. *In situ* AFM measurements of the COM (010) surface indicate that HCA inhibits the growth of hillocks bounded by $\{12\bar{1}\}$ and {021} steps leading to the disappearance of distinct step edges within minutes of introducing the inhibitor (Fig. 1l). Prior *in situ* AFM studies of COM have been reported wherein macromolecular inhibitors, such as proteins[21,23] and glycosaminoglycans[22], have similar effects on surface growth.

When evaluating the lower limits of CA and HCA efficacy by AFM, a unique mode of action compared to conventional mechanisms[14,24,25] of crystal growth inhibition was noted at inhibitor concentrations of $<0.25\,\mu g\,ml^{-1}$. Time-resolved images of COM (100) growth in the presence of CA reveal the appearance of etch pits (Fig. 2a and Supplementary Video 1) that increase in depth *d* with continuous imaging (Fig. 2b). A similar phenomenon was observed for HCA. Periodic snapshots from Supplementary Video 2 presented in Fig. 2c show a hillock on the (100) surface that is initially growing in the absence of inhibitor at supersaturation ratio $S = 4.1$, and is then subjected to the same growth solution containing HCA with a molar ratio $Ca^{2+}/HCA \approx 10^3$. Etch pits immediately form once HCA is introduced into the AFM liquid cell. The etch pits appear to originate at step edges and evolve in both depth and width with imaging time (Extended Data Fig. 4).

To our knowledge, this is the first observation of crystal dissolution in a highly supersaturated growth environment. This effect cannot be attributed to inhibitor complexation with free Ca^{2+} ions in solution, which would require comparable concentrations of inhibitor and solute to reduce calcium concentration below the solubility of COM crystals. Here the inhibitor concentration is nearly three orders of magnitude

[1]Department of Chemical and Biomolecular Engineering, University of Houston, Houston, Texas 77204, USA. [2]Litholink Corporation, Laboratory Corporation of America Holdings, Chicago, Illinois 60612, USA. [3]Department of Chemical Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, USA.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

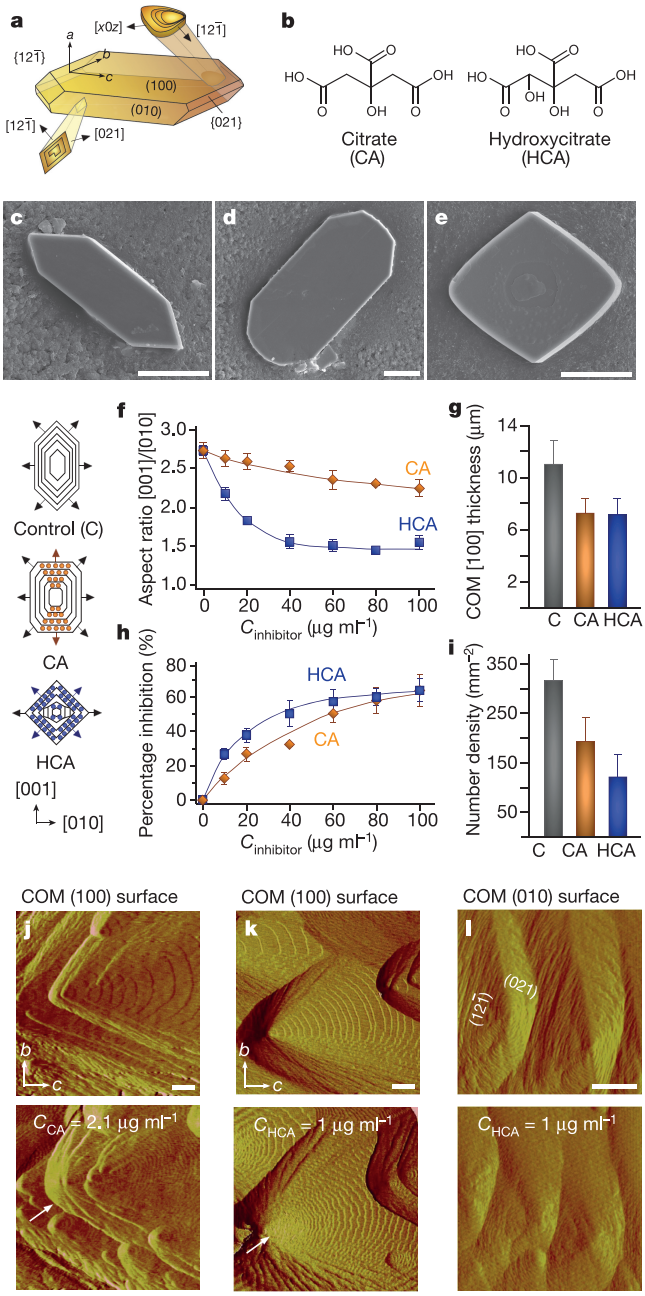


Figure 1 | Effect of inhibitors on COM crystallization. a, COM crystal habit with indexed faces. **b**, Molecular structures of CA and HCA. **c–e**, SEM images of COM crystals in the absence of inhibitor (control C) (**c**) and in the presence of 20 μg ml^{-1} CA (**d**) and in the presence of 20 μg ml^{-1} HCA (**e**). Scale bars, 20 μm. **f**, Changes in COM [001]/[010] (or *c*/*b*) aspect ratio with inhibitor concentration, $C_{inhibitor}$ ($n \geq 150$). The schematics (next to **f** and **h**) depict inhibitor specificity. **g**, Comparison of COM [100] thickness at $C_{inhibitor} = 20$ μg ml^{-1} ($n \geq 150$). **h**, Percentage inhibition of COM crystal growth as a function of $C_{inhibitor}$ ($n \geq 8$). **i**, Comparison of COM crystal number density ρ_{COM} at $C_{inhibitor} = 20$ μg ml^{-1} ($n \geq 3$ batches; see Extended Data Fig. 3). Error bars in **f** and **h** are 2 standard deviations (s.d.); those in **g** and **i** are 1 s.d. **j–l**, AFM deflection mode images of a (100) surface (**j**, **k**) and a (010) surface (**l**) during *in situ* measurements in supersaturated solution ($S = 4.1$). Images of control surfaces (top panels of **j–l**) reveal two-dimensional layer growth. At $C_{CA} = 2.1$ μg ml^{-1} (bottom panel of **j**) we observe jagged step edges, rounding of hillocks, and reduced interstep distances (arrows). Measurements at $C_{HCA} = 1$ μg ml^{-1} (bottom panels of **k** and **l**) indicate the appearance of step protrusions and decreased interstep distances on the (100) face, whereas step edges become indistinguishable on the (010) face. Scale bars, 1 μm.

less than that of calcium, suggesting that the effect is related to specific inhibitor–crystal interactions.

Time-resolved *in situ* AFM images of COM (010) surface growth indicate that dissolution occurs within a narrow range of HCA concentration, C_{HCA}. Inhibitor–crystal interactions reduce step velocity in the [12$\bar{1}$] and [021] directions at $C_{HCA} < 0.08$ μg ml^{-1}. The velocity monotonically decreases with increasing C_{HCA} (Fig. 2d, e) in a manner that suggests that HCA preferentially binds to step sites on the COM (010) surface (Extended Data Fig. 5). Within the range $0.08\,\mu g\,ml^{-1} \leq C_{HCA} \leq 0.15\,\mu g\,ml^{-1}$, step velocities become negative (Fig. 2e) and layers uniformly dissolve. At $C_{HCA} > 0.15$ μg ml^{-1} we begin to observe the disappearance of distinct steps, as indicated in Fig. 1l. Snapshots from Supplementary Video 3 (Fig. 2f) show the explicit effect of HCA on [12$\bar{1}$] and [021] step advancement. Immediately upon introducing a supersaturated solution ($S = 4.1$) with $C_{HCA} = 0.10$ μg ml^{-1} into the AFM liquid cell, steps recede towards the centre of the hillock and etch pits (arrows in Fig. 2f) appear on terraces at later imaging time. As a means of comparison, we also assess COM (010) surface dissolution in an undersaturated solution ($S = 0.5$) without any inhibitor where time-elapsed images from Supplementary Video 4 reveal negative step velocity and layers uniformly receding (Fig. 2g) in a manner that is almost identical to the effect of HCA in supersaturated solution.

Few studies in literature postulate that inhibitors are capable of inducing crystal dissolution in supersaturated solution. Lutsko *et al.*[26] simulated the effect of occluded impurities in crystals (illustrated in Fig. 3a) and showed that negative step velocity can be achieved when the impurity is sufficiently large and reaches a high surface coverage (that is, small separation distances Δx between occluded impurity molecules). An alternative hypothesis for localized (or virtual) dissolution along step edges[13] describes the effect of modifiers on spiral growth originating from screw dislocations. This mode of action leads to a reduced rate of spiral growth (illustrated in Fig. 3b). The reduced velocity of steps on COM (010) surfaces at low inhibitor concentration is qualitatively consistent with this theoretical mechanism; however, the trend deviates from pre-existing models at higher inhibitor concentration.

For COM growth inhibition at this condition, we propose a new mechanism to describe the effect of CA and HCA based on changes in surface energy γ when the inhibitor adsorbs on COM crystal steps (Fig. 3c) or terraces. For cases when inhibitor–crystal interactions are more energetically favourable than solute–crystal interactions, the adsorbed inhibitor imparts an interfacial strain on the crystal lattice. Strain fields are illustrated in Fig. 3c and d with an arbitrary radius of curvature that includes nearest-neighbour interactions between carboxylic acids of the inhibitor and the carboxylic acid of surface oxalates (Ox) formed via a calcium bridge, $_{(inhibitor)}COO^{-}\ldots Ca^{2+}\ldots {}^{-}OOC_{(Ox,\,COM)}$. We postulate that steps dissolve and release Ox and calcium ions into solution to alleviate this strain, thus generating fresh crystal interfaces for additional inhibitor–crystal interactions to perpetuate dissolution, causing steps to recede towards the centre of the hillock (that is, negative step velocity). At higher inhibitor concentration, increased coverage of inhibitor on COM crystal surfaces places the adsorbed molecules in closer vicinity to each other on nearby step sites. This seemingly slows the release of solute from crystal surfaces owing to mass transport limitations, or steric hindrance, wherein inhibitor diffusion on, or desorption from, the COM surface is rate-limiting. The competing effects of inhibitor adsorption/desorption and solute attachment/dissolution produce corrugated steps (illustrated in Fig. 3d), which is consistent with AFM topographical images of COM (010) surfaces at high HCA concentration (see Fig. 1l).

We used density functional theory (DFT) calculations to rationalize the observed experimental trends and shed light on the action of inhibitors on COM crystal growth (see Methods for details). The calculated HCA binding strength on COM (100) and (021) surfaces (Fig. 3e, f) relative to that of an Ox molecule (Extended Data Fig. 6) indicates that HCA–crystal interactions are more energetically favourable. The net

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

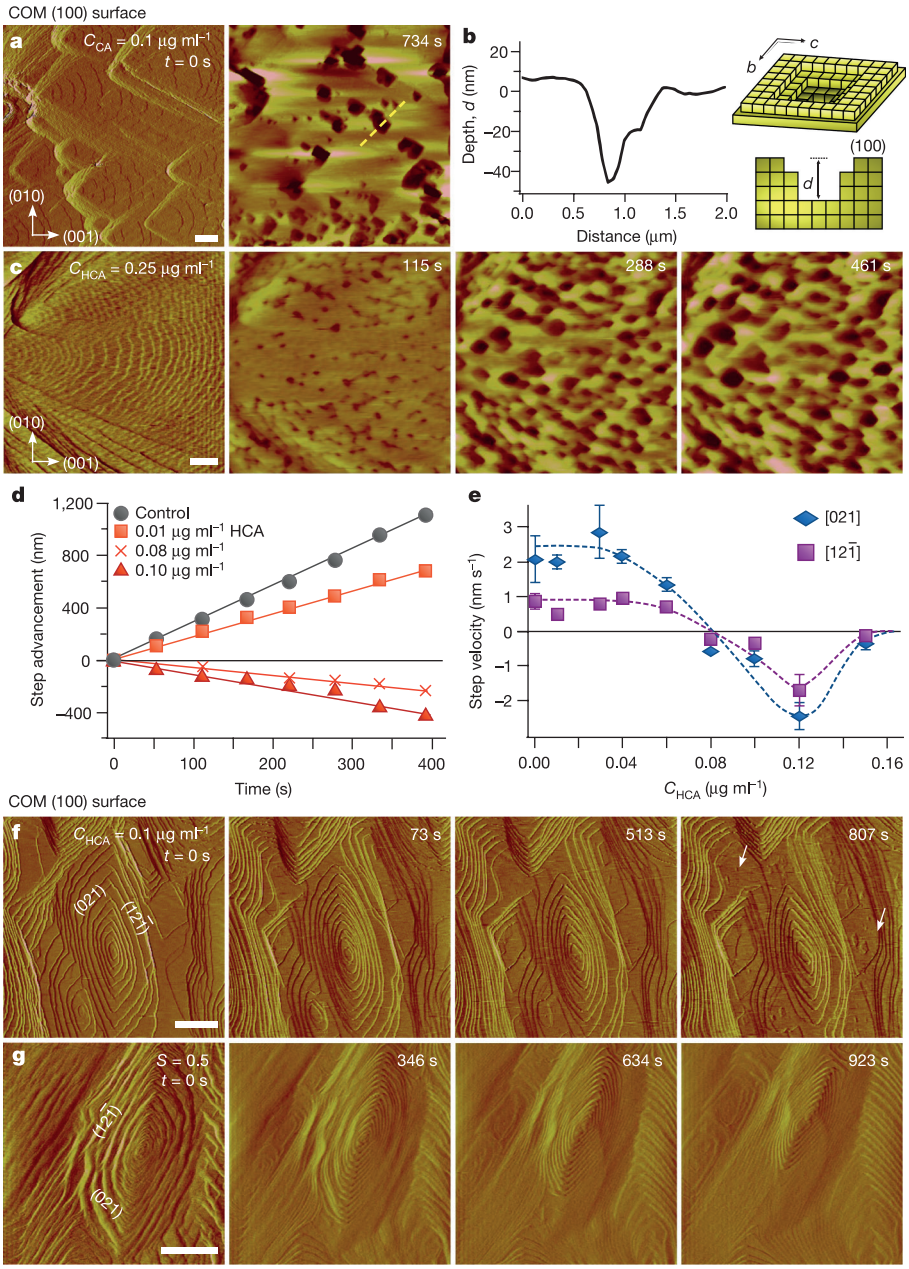


Figure 2 | Time-resolved imaging of COM crystal dissolution. a, *In situ* AFM images of a COM (100) surface growing at 25 °C ($S = 4.1$). Growth solution is continuously delivered to the liquid cell at a rate of 0.2 ml min^{-1} to maintain constant supersaturation. Time-elapsed images from Supplementary Video 1 show etch pit formation on the (100) surface at $C_{CA} = 0.10\,\mu g\ ml^{-1}$. **b**, Depth profile of the etch pit at $t = 734$ s evaluated along the yellow dashed line (in right panel of **a**). **c**, Snapshots from Supplementary Video 2 ($t = 0$–461 s) show etch pit formation on a COM (100) surface at $C_{HCA} = 0.25\,\mu g\ ml^{-1}$. **d**, Measurements of step advancement in the [12$\bar{1}$] direction as a function of imaging time at varying C_{HCA}. Solid lines are linear regression. **e**, Step velocity in the [021] and [12$\bar{1}$] direction monotonically decreases with increased HCA concentration at $C_{HCA} < 0.08\,\mu g\ ml^{-1}$, whereas steps recede (that is, negative velocity) at $C_{HCA} = 0.08$–$0.15\,\mu g\ ml^{-1}$. Dashed lines are interpolated. Error bars are 2 s.d. ($n = 5$). **f**, Time-resolved AFM deflection mode images from Supplementary Video 3 ($t = 0$–807 s) show a receding hillock on a (010) surface at $C_{HCA} = 0.10\,\mu g\ ml^{-1}$. **g**, AFM deflection mode images from Supplementary Video 4 ($t = 0$–923 s) show a (010) hillock dissolving in undersaturated solution ($S = 0.5$) in the absence of inhibitor. Scale bars, 1 μm.

difference between HCA and Ox binding to the COM (100) surface is −31.0 kcal mol^{-1} ($BE_{HCA} = -87.3$ and $BE_{Ox} = -56.3$ kcal mol^{-1}), and the corresponding difference on the (021) surface is −94.7 kcal mol^{-1} ($BE_{HCA} = -170.2$ and $BE_{Ox} = -75.5$ kcal mol^{-1}). The binding preference of HCA for the (021) surface is qualitatively consistent with experimental findings that HCA exhibits a specificity for the apical tips of COM crystals, thus explaining its ability to reduce [021] step velocity. On this basis, we expect that HCA adsorption induces higher strain on the (021) face than the (100) face. Indeed, partial relaxation of COM surfaces in the presence of adsorbed HCA reveals that the (100) face is geometrically unaffected, whereas the (021) face exhibits dislocations (Extended Data Fig. 7), consistent with the proposed mechanism of crystal dissolution in Fig. 3b–d where inhibitor-induced strain alters the γ of crystal faces in a thermodynamically unfavourable manner. To quantify the strain of CA and HCA binding to COM surfaces, we calculated the average displacement δ of atoms in the crystal lattice. Our calculations reveal that inhibitor adsorption on the (100) face has a marginal impact on δ (see Extended Data Table 1); however, HCA binding to the (021) face leads to higher strain ($\delta = 0.71$ Å) compared to CA ($\delta = 0.55$ Å), consistent with HCA's specificity for inhibiting [021] growth. These findings seemingly agree with general observations reported by Wojciechowski and co-workers[27], who showed that strain induced in ductile crystals subjected to constant tensile stress exhibit reduced rates of growth.

Beyond COM surface interactions, we analysed the coordination of organic anions with Ca^{2+} ions in simulations where the number of molecules increases (Fig. 3g–j). Comparison of HCA, CA, and Ox complexation with Ca^{2+} ions (Fig. 3g) reveals binding energies of −174 kcal mol^{-1}, −144 kcal mol^{-1}, and −114 kcal mol^{-1}, respectively. The corresponding number of Ca^{2+} ions complexed per organic anion is 1.5, 1.5, and 1.0. These calculations indicate that HCA and CA display a higher affinity for Ca^{2+} ion complexation relative to Ox. HCA–Ca^{2+} binding is the most energetically favourable, which is attributed to increased hydrogen bonding in the complexes, owing to the presence of an additional hydroxyl group on HCA compared to CA, in conjunction with an observed molecular flexibility of HCA to fold around and protect Ca^{2+} ions (Fig. 3j and Supplementary Video 5). On the basis of these findings, we propose that an inhibitor must satisfy the criterion $BE_{inhibitor\text{-}crystal} \gg BE_{solute\text{-}crystal}$ (or alternatively $BE_{inhibitor\text{-}calcium} \gg BE_{solute\text{-}calcium}$) in order to induce crystal dissolution.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

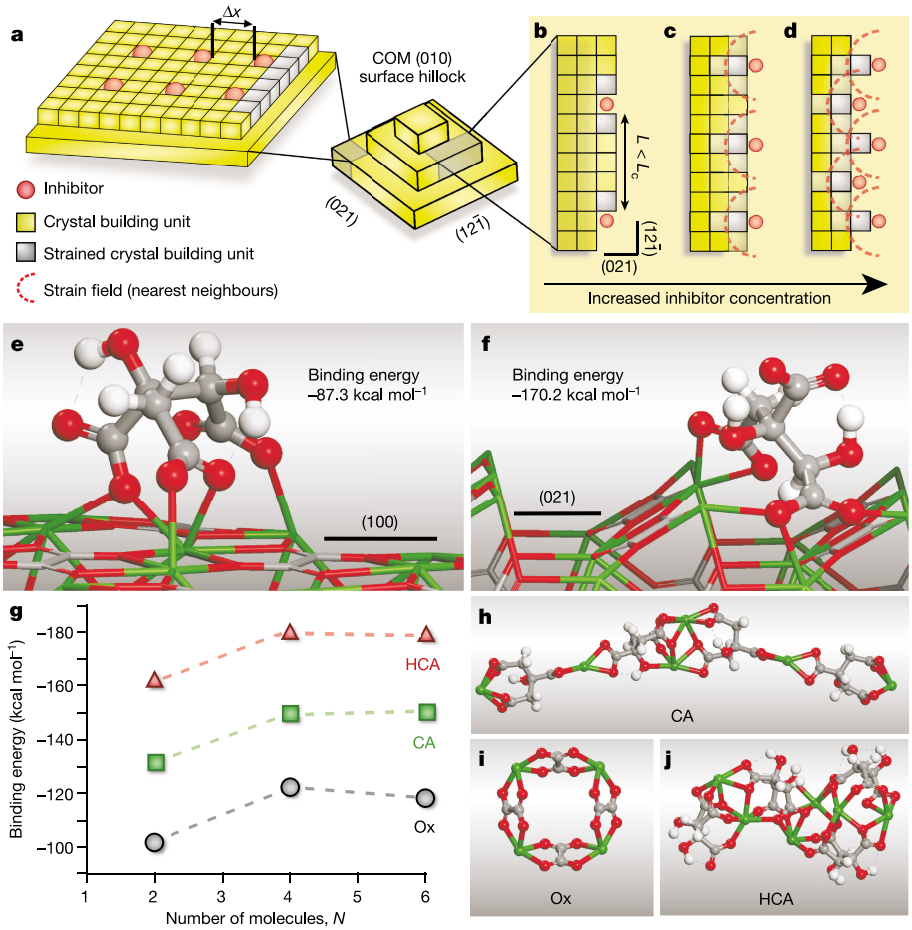


Figure 3 | Mechanism and *in silico* study of inhibitor action. a, Schematic of strain imposed by inhibitors that are incorporated within an advancing, unfinished layer as a result of step pinning. The crystal building unit refers to the solute, either calcium or oxalate. **b**, Idealized mode of step growth inhibition by kink blocking at low inhibitor concentration. **c**, **d**, Mechanism of strain-induced surface dissolution at moderate and high inhibitor concentration, respectively. **e**, **f**, Structural conformation of a single HCA molecule adsorbed on idealized COM (100) and (021) faces (see Extended Data Fig. 7 for results of partially relaxed COM crystal surfaces). Atoms are coloured to represent hydrogen (white), carbon (grey), oxygen (red), and calcium (green). **g**, DFT-calculated binding energy for the complexation of fully dissociated HCA, CA and Ox molecules with calcium ions (see Extended Data Fig. 8 for the calculation of CA with −2 charge). Binding energy data are scaled by the number of molecules N. Dashed lines connecting symbols are added to guide the eye. **h**–**j**, Structures of organic anion and calcium ion complexes (shown for $N = 4$). All binding energy values are exothermic and reported in kcal mol^{-1}.

The relevance of COM in pathological crystallization (kidney stone disease) motivated a detailed study of HCA as a potential replacement for potassium citrate, a supplement that is often prescribed to patients with calcium oxalate kidney stone disease. CA is a normal component of human urine, and is thought to prevent kidney stone formation by complexing Ca^{2+} ions and by acting as an inhibitor of COM crystal growth. Treatment with alkali (most commonly citrate salts) further increases urine CA levels and has been shown to reduce the formation of calcium stones; however, as many as 16% of patients in treatment trials have discontinued the medication owing to side effects[28]. Moreover, the past 30 years has witnessed no major advancement in kidney stone therapy, despite evidence that the incidence rate of kidney stone disease is on the rise[29].

Our findings suggest that HCA has the potential to be an alternative to potassium citrate. We performed an *in vitro* assay to assess the effect of HCA and CA on the upper limit of metastability in urine samples from eight patients with kidney stone disease. This standard assay[30] is used to assess the minimum concentration of Ox required to induce COM nucleation in the presence of inhibitors. Both HCA and CA increased the upper limit of metastability (Fig. 4a) compared to the untreated urine control ($P < 0.02$ for both inhibitors), thus confirming their comparable inhibitory effect on crystal nucleation

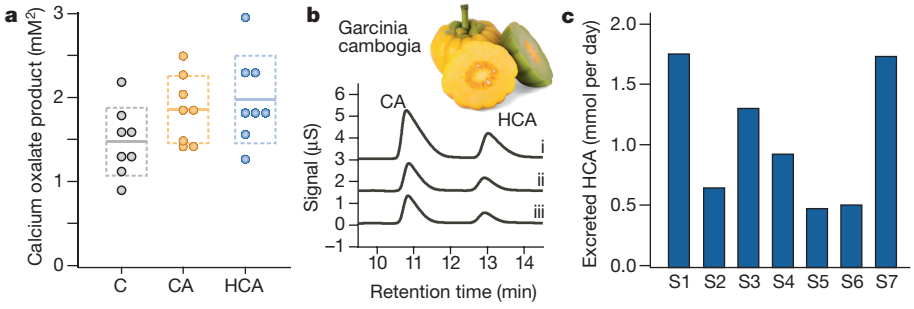


Figure 4 | *In vitro* assay and urinary excretion of HCA. a, The upper limit of metastability in human urine expressed as the calcium oxalate concentration product in the presence of inhibitor ($C_{CA} = C_{HCA} = 2$ mM) exceeds that of the control ($n = 8$ subjects). Solid lines refer to the average and the dashed boxes extend 1 s.d. above/below the average. **b**, Ion chromatography of (i) a standard containing 0.2 mM CA and 0.1 mM HCA, (ii) human urine before treating with isocitrate dehydrogenase, and (iii) human urine from a subject taking HCA supplement after treating with isocitrate dehydrogenase. Data are shifted in the y axis for visual clarity. The inset shows an image of the fruit garcinia cambogia, which contains HCA. **c**, Urinary excretion of HCA from oral administration of clinical-grade garcinia cambogia extract in seven human subjects, S1–S7. Each sample was run in duplicate (data are the averages with coefficients of variation $c_v < 1.5\%$). All subjects tolerated the short-term dosing protocol with no reported side effects.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

in the urine milieu. As some fruits, such as *Garcinia gummi-gutta* (common name garcinia cambogia or Malabar tamarind), contain HCA, we purchased clinical-grade garcinia cambogia extract for human trials to evaluate HCA bioavailability. For HCA to function as a kidney stone treatment, it needs to be excreted in urine. HCA is not a normal component of urine, which we confirmed by measuring HCA concentration in random urine samples in five healthy subjects (three men and two women, mean age 33.4 years) not taking HCA supplement. The result of this control revealed that HCA was below the level of detection (<0.05 mM) in all five subjects. Prior studies have documented detectable blood levels of HCA after ingestion, although only a fraction of ingested drug is absorbed[31,32]. The drug does not appear to be metabolized by humans, but urine excretion rates of HCA have not been reported previously. To this end, we measured urinary excretion of orally ingested HCA in seven non-kidney-stone-forming subjects (five men and two women with mean age 44.6 years old). The subjects took the recommended dose of the supplement, and on the third day of ingestion urine was collected for 24 h and the excreted HCA was measured by ion chromatography (Fig. 4b). As shown in Fig. 4c, the average HCA excretion is 1.1 ± 0.6 mmol per day (with a mean concentration of 0.7 ± 0.6 mM HCA).

To summarize, AFM measurements and DFT calculations reveal a new thermodynamic mechanism of crystal growth inhibition that deviates from classical kinetic models of inhibitor–crystal interactions. We show that two organic anions with high affinity for binding to COM surfaces induce localized strain on the crystal lattice. When adsorbed at moderate coverage, these inhibitors cause COM crystal surfaces to dissolve in supersaturated solution. Our proposed criterion based on the relative binding energies of inhibitor and solute with crystal surfaces may prove to be relevant for predicting the dissolution of other crystalline materials. The comparison of HCA and CA in this study also highlights the subtle nuances of rational design wherein a small difference in molecular structure (for example, insertion of one alcohol group) can substantially alter inhibitor specificity and efficacy. Moreover, we report that HCA shows promise as a potential therapy to prevent kidney stones. HCA may be preferred as a therapy over potassium citrate, such as in patients with alkaline urine where a further increase in urine pH could cause calcium phosphate stones[33]; however, better understanding of HCA metabolism in humans, optimal dosing regimens, and long term safety and tolerability are needed before HCA could be studied in a prospective clinical trial of kidney stone prevention.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 January; accepted 20 June 2016.
Published online 8 August 2016.

1. Evans, J. S. "Tuning in" to mollusk shell nacre- and prismatic-associated protein terminal sequences. Implications for biomineralization and the construction of high performance inorganic-organic composites. *Chem. Rev.* **108,** 4455–4462 (2008).
2. Falini, G., Albeck, S., Weiner, S. & Addadi, L. Control of aragonite or calcite polymorphism by mollusk shell macromolecules. *Science* **271,** 67–69 (1996).
3. Meldrum, F. C. Calcium carbonate in biomineralisation and biomimetic chemistry. *Int. Mater. Rev.* **48,** 187–224 (2003).
4. Dunlop, J. W. C. & Fratzl, P. Biological composites. *Annu. Rev. Mater. Res.* **40,** 1–24 (2010).
5. Killian, C. E. *et al.* Mechanism of calcite co-orientation in the sea urchin tooth. *J. Am. Chem. Soc.* **131,** 18404–18409 (2009).
6. Young, J. R., Didymus, J. M., Bown, P. R., Prins, B. & Mann, S. Crystal assembly and phylogenetic evolution in heterococcoliths. *Nature* **356,** 516–518 (1992).
7. Rimer, J. D. *et al.* Crystal growth inhibitors for the prevention of L-cystine kidney stones through molecular design. *Science* **330,** 337–341 (2010).
8. Olafson, K. N., Ketchum, M. A., Rimer, J. D. & Vekilov, P. G. Mechanisms of hematin crystallization and inhibition by the antimalarial drug chloroquine. *Proc. Natl Acad. Sci. USA* **112,** 4946–4951 (2015).
9. Weissbuch, I. & Leiserowitz, L. Interplay between malaria, crystalline hemozoin formation, and antimalarial drug action and design. *Chem. Rev.* **108,** 4899–4914 (2008).
10. Orme, C. A. *et al.* Formation of chiral morphologies through selective binding of amino acids to calcite surface steps. *Nature* **411,** 775–779 (2001).
11. Davis, K. J., Dove, P. M. & De Yoreo, J. J. The role of Mg^{2+} as an impurity in calcite growth. *Science* **290,** 1134–1137 (2000).
12. Graether, S. P. *et al.* β-helix structure and ice-binding properties of a hyperactive antifreeze protein from an insect. *Nature* **406,** 325–328 (2000).
13. Sizemore, J. P. & Doherty, M. F. A new model for the effect of molecular imposters on the shape of faceted molecular crystals. *Cryst. Growth Des.* **9,** 2637–2645 (2009).
14. Weissbuch, I., Addadi, L., Lahav, M. & Leiserowitz, L. Molecular recognition at crystal interfaces. *Science* **253,** 637–645 (1991).
15. Dey, A., de With, G. & Sommerdijk, N. In situ techniques in biomimetic mineralization studies of calcium carbonate. *Chem. Soc. Rev.* **39,** 397–409 (2010).
16. Coe, F. L., Parks, J. H. & Asplin, J. R. Medical process—the pathogenesis and treatment of kidney-stones. *N. Engl. J. Med.* **327,** 1141–1152 (1992).
17. Wesson, J. A. & Ward, M. D. Pathological biomineralization of kidney stones. *Elements* **3,** 415–421 (2007).
18. Nancollas, G. H. & Gardner, G. L. Kinetics of crystal growth of calcium oxalate monohydrate. *J. Cryst. Growth* **21,** 267–276 (1974).
19. Ryall, R. L., Harnett, R. M. & Marshall, V. R. The effect of urine, pyrophosphate, citrate, magnesium and glycosaminoglycans on the growth and aggregation of calcium-oxalate crystals in vitro. *Clin. Chim. Acta* **112,** 349–356 (1981).
20. Phillips, R. *et al.* Citrate salts for preventing and treating calcium containing kidney stones in adults. *Cochrane Database Syst. Rev.* **10,** CD010057 (2015).
21. Qiu, S. R. *et al.* Molecular modulation of calcium oxalate crystallization by osteopontin and citrate. *Proc. Natl Acad. Sci. USA* **101,** 1811–1815 (2004).
22. Farmanesh, S. *et al.* Specificity of growth inhibitors and their cooperative effects in calcium oxalate monohydrate crystallization. *J. Am. Chem. Soc.* **136,** 367–376 (2014).
23. Wang, L. & Nancollas, G. H. Calcium orthophosphates: crystallization and dissolution. *Chem. Rev.* **108,** 4628–4669 (2008).
24. Chernov, A. A. Formation of crystals in solutions. *Contemp. Phys.* **30,** 251–276 (1989).
25. De Yoreo, J. J. & Vekilov, P. G. in *Biomineralization* Vol. 54 (eds Dove, P. M., De Yoreo, J. J. & Weiner, S.) 57–93 (Mineralogical Society of America, 2003).
26. Lutsko, J. F. *et al.* Crystal growth cessation revisited: the physical basis of step pinning. *Cryst. Growth Des.* **14,** 6129–6134 (2014).
27. Ristić, R. I., Sherwood, J. N. & Wojciechowski, K. Assessment of the strain in small sodium-chlorate crystals and its relation to growth rate dispersion. *J. Cryst. Growth* **91,** 163–168 (1988).
28. Ettinger, B. *et al.* Potassium-magnesium citrate is an effective prophylaxis against recurrent calcium oxalate nephrolithiasis. *J. Urol.* **158,** 2069–2073 (1997).
29. Scales, C. D. J., Smith, A. C., Hanley, J. M. & Saigal, C. S. Prevalence of kidney stones in the United States. *Eur. Urol.* **62,** 160–165 (2012).
30. Asplin, J. R. *et al.* Reduced crystallization inhibition by urine from men with nephrolithiasis. *Kidney Int.* **56,** 1505–1516 (1999).
31. van Loon, L. J. C. *et al.* Effects of acute (-)-hydroxycitrate supplementation on substrate metabolism at rest and during exercise in humans. *Am. J. Clin. Nutr.* **72,** 1445–1450 (2000).
32. Loe, Y. C. C., Bergeron, N., Rodriguez, N. & Schwarz, J. M. Gas chromatography/mass spectrometry method to quantify blood hydroxycitrate concentration. *Anal. Biochem.* **292,** 148–154 (2001).
33. Parks, J. H., Worcester, E. M., Coe, F. L., Evan, A. P. & Lingeman, J. E. Clinical implications of abundant calcium phosphate in routinely analyzed kidney stones. *Kidney Int.* **66,** 777–785 (2004).

Supplementary Information is available in the online version of the paper.

Acknowledgements J.D.R. acknowledges support from the National Science Foundation (grant 1207441) and the Welch Foundation (grant E-1794). G.M. acknowledges start-up funds from the University of Pittsburgh and computational support from the Center for Simulation and Modeling, and the Extreme Science and Engineering Discovery Environment, which is supported by the National Science Foundation (grant ACI-1053575).

Author Contributions J.C. performed data collection and analysis for bulk crystallization and *in situ* AFM studies, I.G. performed *in vitro* experiments in urine and analysed human trial samples, and M.G.T. performed DFT calculations. J.D.R. wrote the paper with help from G.M. and J.R.A., with all three authors contributing to the design and analysis of experiments. I.G. and M.G.T. contributed equally. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.D.R. (jrimer@central.uh.edu) and J.R.A. (asplinj@labcorp.com).

Reviewer Information *Nature* thanks J. Lieske, M. Sleutel and the other anonymous reviewer(s) for their contribution to the peer review of this work.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

Materials. The following reagents were purchased from Sigma Aldrich (St Louis, MO): calcium chloride dihydrate (ACS Reagent, 99+%), sodium oxalate ($Na_2C_2O_4$, >99%), sodium hydroxide (98%), hydrochloric acid (37%), sodium citrate tribasic dihydrate (ACS reagent, ≥99.0%), and potassium hydroxycitrate tribasic monohydrate (95%). Sodium chloride (99.9% ultrapure bioreagent) was purchased from JT Baker. Deionized water used in all experiments was purified with an Aqua Solutions RODI-C-12A purification system (18.2 MΩ). All reagents were used as received without further purification.

Garcinia cambogia was purchased from Swanson Health Products (Super CitriMax Clinical Strength Garcinia Cambogia Extract, item SWD051). The recommended serving size (2 capsules) contained the following components and corresponding mass per serving: calcium (120 mg), chromium (130 μg), potassium (180 mg), and garcinia cambogia extract (1.5 g) containing HCA (900 mg).

COM bulk crystallization. Batch crystallization was carried out in a 20-ml glass vial by dissolving NaCl in deionized water, then adding 0.7 ml of 10 mM $CaCl_2$ stock solution. A clean glass slide (about 1.3 × 1.3 cm^2) was placed at the bottom of the vial to collect the crystals for microscopy. The sample vial was then placed in an oven set to 60 °C for 1 h to ensure the solution reached the set point temperature for crystallization. Subsequently, 0.7 ml of 10 mM $Na_2C_2O_4$ stock solution was added to the vial dropwise while continuously stirring at a rate of about 400 r.p.m. To investigate the effect of growth inhibitors (that is, CA or HCA) on COM crystallization, an appropriate quantity of the inhibitor was added to the growth solution before $Na_2C_2O_4$ addition. The final growth solution had a composition of 0.7 mM $CaCl_2$:0.7 mM $Na_2C_2O_4$:150 mM NaCl:x μg ml^{-1} inhibitor (where x = 0–100) and a total volume of about 10 ml. Crystallization was performed at 60 °C for 3 days at static conditions (that is, without stirring or agitation). The glass slide (substrate) was removed from the solution, gently washed with deionized water, and dried at room temperature before analysis. The pH of COM growth solution was measured before and after crystallization using an Orion 3-Star Plus pH benchtop meter with a ROSS Ultra electrode (8102BNUWP).

Characterization of COM crystallization. The size and morphology of COM crystals prepared in the absence and presence of inhibitors was assessed by optical microscopy using a Leica DM2500-M instrument. Brightfield images were obtained in reflectance mode to quantify the crystal dimensions, which we report as length L in the [001] direction, width W in the [010] direction, and the length-to-width (L/W) aspect ratio. A minimum of 150 crystals from three separate batches were measured to obtain an average aspect ratio for data reported in Fig. 1f. COM crystal number density, ρ_{COM}, was measured as the number of crystals per area of glass slide. The ρ_{COM} values reported in Fig. 1i are an average of at least ten areas on glass slides from three separate batches (each micrograph area is 647 μm × 484 μm). The COM [100] thickness was measured from a combination of optical and electron micrographs. Scanning electron microscopy (SEM) was performed using a FEI 235 Dual-Beam Focused Ion-beam instrument equipped with a SEM sample extraction probe. SEM samples were prepared by gently pressing COM crystals on the glass slide to carbon tape to transfer crystals to the sample disk. Each sample was coated with a layer of carbon (about 20 nm thick) to reduce the effects of electron beam charging. The average values reported in Fig. 1g were obtained from measurements of more than 150 crystals from three separate batches.

The effect of growth inhibitors on the kinetics of COM crystallization was measured using a calcium ISE from ThermoScientific equipped with an Orion 9720BNWP ionplus electrode. ISE measurements track the temporal reduction in free calcium ion concentration in the growth solution during crystallization (including the effects of both nucleation and crystal growth)[34]. Growth solutions were prepared similar to COM bulk crystallization, but at room temperature using a solution with supersaturation ratio S = 3.8 and a composition of 0.5 mM $CaCl_2$:0.5 mM $Na_2C_2O_4$:150 mM NaCl:x μg ml^{-1} inhibitor (where x = 0–100). For ISE studies, we used the calcium oxalate solubility product reported by ref. 35 in order to calculate the calcium oxalate supersaturation. ISE measurements were performed at room temperature under constant stirring (about 1,200 r.p.m.) to minimize the induction time[22,34]. Plots of consumed calcium ion concentration as a function of time were generated for each inhibitor concentration. A minimum of eight measurements were performed for each data point reported in Fig. 1h. The data were normalized by subtracting the concentration of the initial time point (see Extended Data Fig. 1). The approximate linear slope of these curves during the first 40 min of crystallization corresponds to the rate of Ca^{2+} depletion. The efficacy of growth inhibitors was determined by the percentage inhibition, which was calculated by comparing the change in slope of the growth curve in the presence of inhibitor relative to that in the absence of inhibitor (that is, the control). Prior to ISE measurements, the electrode was calibrated using a standard calcium solution (0.1 M, Orion Ion Plus), which was diluted with deionized water to three concentrations: 0.1 mM, 1.0 mM and 10.0 mM. The ionic strength of each solution was adjusted using a standard solution (ISA, Thermo Scientific), which was added in a 1:50 volume ratio of ISA-to-standard.

In situ AFM was performed using a Digital Instruments Multimode IV (Santa Barbara, CA) to examine topographical images of COM crystals and capture the dynamics of surface growth in real time. COM crystals (about 50 μm in length) were mounted on an AFM specimen disk (Ted Pella) covered with a thin layer of thermally curable epoxy (MasterBond EP21AOLV), in accordance with previously reported protocols[22,36,37]. The epoxy was partially cured in an oven for about 20 min at 60 °C. COM crystals on glass slides from bulk crystallization assays (in the absence of inhibitor) were immobilized on the partially cured epoxy by gently pressing the glass slide to transfer crystals with either their (100) or (010) faces oriented normal to the specimen surface. The sample was then placed in an oven at 60 °C for an additional 3 h to completely cure the epoxy. All AFM measurements were performed using silicon nitride probes with gold reflex coating and a spring constant of 0.15 N m^{-1} (Olympus, TR800PSA). *In situ* experiments were performed to monitor surface growth in supersaturated calcium oxalate solution. We measured the velocity of step advancement and changes in hillock morphology on COM (010) and (100) surfaces in the absence or presence of inhibitors. The reported step velocity (Fig. 2e) is the average of at least 5 measurements of different steps. For *in situ* AFM measurements, a growth solution with supersaturation ratio S = 4.1 was prepared, similar to the solution used for ISE measurements, but with a composition of 0.18 mM $CaCl_2$:0.18 mM $Na_2C_2O_4$:x μg ml^{-1} inhibitor (where x = 0–2.1). The inhibitor concentration used in AFM measurements is lower than that employed in bulk crystallization[22,37] owing to fewer crystals (that is, smaller COM surface area) on the AFM specimen disk. The AFM instrument was equipped with a fluid cell (model MTFML) containing two ports for inlet and outlet flow to maintain constant supersaturation during continuous imaging. The growth solution was delivered to the liquid cell using a dual syringe pump (CHEMYX, Fusion 200) with an in-line mixing configuration[38] to combine $CaCl_2$ and $Na_2C_2O_4$ solutions with a combined flow rate of 0.2 ml min^{-1}. Inhibitors were introduced into the $Na_2C_2O_4$ solution at the appropriate concentration (taking into account the effect of dilution at the in-line flow connection). Continuous imaging was performed in contact mode with a scan rate of 7.0–9.2 Hz at 256 lines per scan.

***In vitro* assays of COM crystallization in human urine.** The upper limit of metastability was measured using a modified version of the method described by ref. 30. Urine aliquots were obtained over a 24-h urine collection period from eight patients with kidney stones. All urine samples were brought to a pH of 5.7 by the addition of potassium hydroxide or HCl as needed. Each urine sample was studied with no additive or with either CA or HCA added to increase their concentration by 2 mM. For each sample, 200 μl of urine was added to 12 wells of a 96-well microlitre plate. Solutions of increasing concentration of Ox were pipetted into the urine aliquots in the wells. The plate was placed on a shaker for 3 h at 37 °C and then the turbidity of solutions in each well were measured at 620 nm wavelength using a VMax kinetic ELISA microplate reader (Molecular Devices, Sunnyvale, CA). The well at which turbidity increased determined the point of crystallization and the Ox concentration at this point is the amount of Ox in the urine measured at baseline plus the amount of Ox added to the urine in the well showing increased turbidity. The results are presented as the calcium oxalate concentration product (used as a surrogate of supersaturation) at the point of crystallization (see Fig. 4a). In addition to their crystal inhibition activity, both CA and HCA complex calcium in solution, lowering the concentration of ionized calcium. Both mechanisms should contribute to changes in the upper limit of metastability relative to the control, indicative of an inhibition of nucleation. Each urine sample was run in duplicate and the results were averaged. Statistical comparison was performed using the non-parametric Wilcoxon test.

Human trials of HCA bioavailability. There were no prior reports of measurements of HCA in human urine in people not consuming HCA supplement. We tested the hypothesis that HCA is not a normal constituent of human urine by measuring the concentration of HCA in random urine samples from five healthy subjects (protocol number 1061857 Western Institutional Review Board).

The hypothesis of the human trial study was that HCA, when orally administered, will be excreted in urine. To test this hypothesis, we assessed urinary excretion to confirm the bioavailability of HCA through oral administration. The protocol was approved by the University of Houston Internal Review Board (case 15176-01). Recruitment was limited to subjects between 21 and 65 years of age. Pregnant women and subjects with known severe chronic kidney disease (stage 4 or 5) were excluded. All samples were collected and analysed with informed consent.

The supplement used in the human trial was Super CitriMax Clinical Strength Garcinia Cambogia Extract. The active ingredient, HCA, is an inhibitor of ATP citrate lyase and is presumed to reduce lipogenesis as its mechanism of action for inducing weight loss[31]. Each serving (2 capsules) contained 1.5 g garcinia cambogia extract with 900 mg active ingredient. The subjects were asked to take garcinia

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

cambogia extract for three days at the dose recommended by the manufacturer (that is, two capsules three times a day). On the third day of garcinia cambogia treatment, the subjects collected urine for 24 h. The urine was collected unrefrigerated using an antimicrobial preservative.

HCA concentration was measured by ion chromatography using an ICS-2000 system (Dionex Corp., Sunnyvale, CA) with AS11 guard and analytic columns, potassium hydroxide eluent, and a conductivity detection system. Because isocitrate co-elutes with HCA on this system, urine samples were pre-treated with isocitrate dehydrogenase to remove isocitrate interference. Hydroxycitric acid calcium salt, (−)-(P), purchased from ChromaDex Inc. (Irvine, CA) was used as a standard.

DFT calculations. Molecular orbital DFT calculations were performed using the Turbomole/6.6 program package[39]. We used the BP86 functional[40,41] and accounted for dispersion energy corrections through the D3 method[42] appropriate for capturing hydrogen bonds originating by the presence of HCA and CA inhibitors. The BP86 method has been shown to successfully capture the aggregation behaviour of metal–cation complexed organic acids[43]. The resolution of identity[44] approximation along with multipole accelerated resolution of indices (MARI-J)[45] were used to accelerate the calculations. We used the def2-SV(P) basis set[46] and accounted for solvent effects through the COSMO continuum solvation model (the solvent was water; dielectric constant $\varepsilon = 78.46$)[47]. The COM nanoparticle with a (100) surface termination consisted of 168 atoms ($Ca_{24}C_{48}O_{96}$), whereas the one with a (021) termination consisted of 133 atoms ($Ca_{19}C_{38}O_{76}$). The (001)-to-(100) step consisted of 224 atoms ($Ca_{32}C_{64}O_{128}$). We have used the whewellite (monoclinic) crystal unit cell[48] to build the COM nanoparticles (with H_2O molecules being removed and simulated by implicit solvent effects). The (100) and (021) surfaces and (001)-to-(100) step were kept frozen and the inhibitors were allowed to fully relax in our calculations. We also performed calculations where the surface of COM was allowed to relax and the overall adsorption trends remained the same (see partial relaxation calculations in Extended Data Fig. 7). The isomer of HCA in garcinia cambogia was taken into account, consistent with the natural extract used in human trials[49]. The binding energy of inhibitors and oxalate to crystal surfaces is defined as:

$$\mathrm{BE} = E_{\mathrm{inhibitor+COM}} - E_{\mathrm{inhibitor}} - E_{\mathrm{COM}} \quad (1)$$

where E_x represents the total electronic energy of species x. For calculations of the 'inhibitor + COM' species, the deprotonated forms of the acids were placed on the COM surfaces and their hydrogens (from the deprotonation) were placed at a position far away from the interaction centre and anti-diametric on the COM nanoparticle to bring the system to a neutral charge state and avoid calculations under high charge states (that is, charge counterbalance). The energy of the inhibitor in the gas phase, labelled as $E_{\mathrm{inhibitor}}$, corresponds to molecules in their protonated (neutral) state. Analogously to the previous expression, the binding energy of the complexes used in determining the affinity of HCA, CA and Ox for Ca^{2+} is defined as:

$$\mathrm{BE}_{\mathrm{complex}} = E_{\mathrm{complex}} + n\frac{d}{2}E_{\mathrm{H_2}} - n\frac{d}{2}E_{\mathrm{Ca}} - nE_{\mathrm{inhibitor}} \quad (2)$$

where n represents the number of organic molecules in the complex, d is the deprotonation state of the inhibitor (that is, $d = 1$ corresponds to single, $d = 2$ to double, and $d = 3$ to triple deprotonated states, equivalently), and E_x represents the electronic energy of species x. The inhibitor is in the protonated form in the gas phase and we use H_2 as a reference state for the hydrogen of the acids. Approximately four different initial conformations were taken into consideration for each inhibitor–surface and complexation calculation, wherein we report the lowest-energy conformations. Our obtained minima from the optimization calculations were further validated by the absence of any imaginary frequencies on the complexes and on the inhibitors interacting with COM crystal surfaces.

To quantify crystal lattice strain, we employ a geometric comparison between the frozen and relaxed structures of COM surfaces in the presence and absence of adsorbates (growth inhibitors). We developed a distance metric to quantify the displacement of relaxed atoms (relative to their frozen state) on COM crystal surfaces by taking the average absolute value of the (x, y, z) coordinate displacements. The average displacement δ is represented as:

$$\delta = \frac{\sum_{i=1}^{N_{\mathrm{atoms}}} \sqrt{(x_{i,\mathrm{frozen}} - x_{i,\mathrm{relaxed}})^2 + (y_{i,\mathrm{frozen}} - y_{i,\mathrm{relaxed}})^2 + (z_{i,\mathrm{frozen}} - z_{i,\mathrm{relaxed}})^2}}{N_{\mathrm{atoms}}} \quad (3)$$

where N_{atoms} represents the number of atoms that are relaxed on the surface of the COM crystallographic plane.

CA and HCA speciation. The pH of crystallization media impacts the net charge of the inhibitor. Both CA and HCA have three dissociation constants corresponding to each of their carboxylic acid groups. In physiologically relevant environments, such as the kidney, the pH of urine varies between 5 and 8 (ref. 50). *In vitro* COM crystallization assays are performed at approximately neutral pH. For instance, the growth solutions employed in this study have $\mathrm{pH} = 6.2 \pm 0.2$, which is identical to previous *in vitro* assays published by Rimer and co-workers[22,34,37,51]. As shown in Extended Data Fig. 2a, equilibrium calculations at pH 6.2 predict HCA to be predominantly in the fully dissociated state (that is, net charge = −3) while approximately 40% of CA species are in the fully dissociated state, labelled as CA^{3-} or $C_6H_5O_7^{3-}$.

The acid–base speciation reactions for CA along with the respective dissociation constants, $\mathrm{p}K_{\mathrm{a}_i}$, are the following:

$$C_6H_8O_7 \xleftrightarrow{\mathrm{p}K_{\mathrm{a}_1}=3.13} C_6H_7O_7^- + H^+ \quad (4)$$

$$C_6H_7O_7^- \xleftrightarrow{\mathrm{p}K_{\mathrm{a}_2}=4.76} C_6H_6O_7^{2-} + H^+ \quad (5)$$

$$C_6H_6O_7^{2-} \xleftrightarrow{\mathrm{p}K_{\mathrm{a}_3}=6.40} C_6H_5O_7^{3-} + H^+ \quad (6)$$

The reported $\mathrm{p}K_{\mathrm{a}}$ values were obtained from Martell and Smith[52]. There are discrepancies in the reported $\mathrm{p}K_{\mathrm{a}_3}$ value owing to the ionic strength dependency of the dissociation constant. For instance, some references report $\mathrm{p}K_{\mathrm{a}_3} = 5.66$ when the ionic strength is 100 mM (ref. 52), which is less than the ionic strength of calcium oxalate solutions used in ISE and bulk crystallization assays. On this basis, it is likely that the dominant species in calcium oxalate growth solutions is CA^{3-}. DFT calculations of CA–crystal interactions and CA–calcium ion complexes in the manuscript were performed using CA^{3-}; however, equivalent DFT calculations were performed with CA^{2-}. The results of these calculations, which are presented in Extended Data Figs 6 and 8, reveal that the conclusions reached in the manuscript are not altered by CA charge. Moreover, we conducted bulk crystallization, ISE, and *in situ* AFM measurements in growth solutions at pH 6.2 (nominal condition) and pH 8.0 to assess the influence of CA charge on its efficacy as a COM crystal inhibitor. At pH 8.0, approximately 99.8% of CA species are in the fully dissociated state (neglecting the effect of ionic strength). Bulk crystallization assays reveal approximately no change in crystal morphology and size at these two pH values. Similarly, *in situ* AFM measurements at pH 6.2 and 8.0 at $C_{\mathrm{CA}} = 0.1\,\mu\mathrm{g\,ml^{-1}}$ showed no apparent change in etch pit formation on COM (100) surfaces. ISE experiments at pH 6.2 and 8.0 reveal subtle differences in the percent inhibition of COM growth at $C_{\mathrm{CA}} = C_{\mathrm{HCA}} = 20\,\mu\mathrm{g\,ml^{-1}}$ (Extended Data Fig. 2b). The percent inhibition of COM growth is slightly reduced at higher pH, but HCA is the most effective inhibitor irrespective of solution alkalinity.

The speciation reactions and corresponding equilibrium dissociation constants for HCA are the following[53]:

$$C_6H_8O_8 \xleftrightarrow{\mathrm{p}K_{\mathrm{a}_1}=2.90} C_6H_7O_8^- + H^+ \quad (7)$$

$$C_6H_7O_8^- \xleftrightarrow{\mathrm{p}K_{\mathrm{a}_2}=4.29} C_6H_6O_8^{2-} + H^+ \quad (8)$$

$$C_6H_6O_8^{2-} \xleftrightarrow{\mathrm{p}K_{\mathrm{a}_3}=5.11} C_6H_5O_8^{3-} + H^+ \quad (9)$$

As shown in Extended Data Fig. 2a, the percentage of fully dissociated HCA (labelled as HCA^{3-} or $C_6H_5O_8^{3-}$) at pH 6.2 is 92%. At pH 8.0 (that is, the upper limit of urine), the percentage of HCA^{3-} is nearly 100%. As such, DFT calculations in the manuscript were performed using the dominant species, HCA^{3-}.

COM (100) surface dissolution in the presence of CA and HCA. In Extended Data Fig. 4 we provide a detailed analysis of etch pit formation on a COM (100) surface in the presence of $C_{\mathrm{HCA}} = 0.25\,\mu\mathrm{g\,ml^{-1}}$. Time-resolved *in situ* AFM images at periodic times reveal the formation and growth of etch pits (Extended Data Fig. 4a–c). For each dashed line in the AFM image, the corresponding height profiles are provided in Extended Data Fig. 4d–f, respectively. The nominal growth of hillocks on the (100) surface in the absence of inhibitor (Extended Data Fig. 4d) occurs by the advancement of single steps. Each step has an average height of 0.4 nm, which is the approximate size of the unit cell parameter in the a direction. Upon the addition of HCA, etch pits monotonically evolve in both depth (Extended Data Fig. 4g) and width (Extended Data Fig. 4h) with imaging time. The concentration of inhibitor in this study is approximately three orders of magnitude less than the concentration of Ca^{2+} ions in supersaturated solution, indicating the effect of etch pit formation is not solely attributed to reduced calcium oxalate supersaturation as a result of inhibitor–calcium ion complexation.

Mechanism of HCA-induced dissolution of COM crystals. Common mechanisms of crystal growth inhibition include step pinning and kink blocking[24,54]. The former involves the adsorption of inhibitors on terraces or step edges, which impede the advancement of steps. Inhibitors adsorbed with their average

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

adsorbate-to-adsorbate spacing comparable to the critical radius of curvature r_c for the step (that is, high surface coverage at high inhibitor concentration) impedes step advancement, leading to reduced step velocity and ultimately suppressed growth when the step radius approaches r_c (refs 54 and 55). The mode of action for HCA on COM crystallization at low inhibitor concentration, as inferred from time-resolved AFM images of the COM (010) surface, does not appear to be step pinning owing to the absence of protrusions on steps (which is a characteristic trait of inhibitors that operate by a step-pinning mode of action[54]). AFM measurements of COM (100) surface growth at high concentration of either CA or HCA reveal the formation of irregular-shaped steps, which may indicate that inhibitors bind to step edges on this surface (consistent with a previously proposed mechanism for CA inhibition[21]).

Kink blocking is commonly observed when the crystal grows by a screw dislocation mechanism[56,57] wherein inhibitors adsorb to kink sites and reduce the kink density and extend the critical length of the step[13]. The net result is a decreased rate of step advancement within the plane of measurement, as well as reduced growth of hillocks in the direction normal to this plane. AFM studies of COM growth have shown that steps emanating from screw dislocations on the (100) and (010) surfaces advance across the crystal plane. Burton, Cabrera and Frank[56] derived a theoretical model of spiral growth predicting the rate of crystal growth normal to the basal face, $G_{[hkl]}$, as follows:

$$G_{[100]} = \frac{(v_i)_{hkl} h_{hkl}}{(y_i)_{hkl}} = \frac{h}{\tau} \quad (10)$$

where h_{hkl} is the height of (hkl) steps advancing along the COM (100) plane, y_i is the interstep distance, and v_i is the velocity of step advancement[58–60]. The subscript refers to the ith edge of growth hillocks, which advance across the surface in spiral patterns with a characteristic rotation time τ. Observations of decreased COM [100] thickness in Fig. 1g are qualitatively consistent with this theoretical equation.

It is energetically more favourable for inhibitors to adsorb on kink sites rather than on step edges. Both modes of action can alter the shape of crystals through specific inhibitor–step or inhibitor–kink interactions. In our studies of COM crystallization, the mode of action for CA and HCA at low inhibitor concentration inferred from *in situ* AFM images does not appear to be kink blocking. To quantitatively analyse AFM data, we constructed Bliznakov plots, which represent the relationship between relative step velocity and inhibitor concentration. In Extended Data Fig. 5a we plot a hypothetical v/v_0 trend with increasing inhibitor concentration as a function of calcium oxalate supersaturation where v and v_0 are step velocities in the presence and absence of inhibitor, respectively. For the step-pinning mode of action, this plot should exhibit a monotonic decrease in v/v_0 with increasing inhibitor concentration (see ref. 8 for example). Plots for HCA (Extended Data Fig. 5b) deviate from this trend, suggesting that HCA does not bind to the (010) surface. For inhibitors that operate in a kink-blocking mode of action, a plot of $v_0/(v_0 - v)$ with increasing inverse inhibitor concentration should produce a linear trend; however, plots for HCA exhibit no apparent trend, which leads us to believe that HCA preferentially binds to steps, as illustrated in Fig. 3c. Note that the plateau region for v (Fig. 2e) observed at low inhibitor concentration ($C_{HCA} < 0.04\,\mu g\ ml^{-1}$) has also been reported by others examining the effects of proteins and polyamino acids on COM (010) and (100) step advancement[38]. In these studies, it was suggested that the size of the step relative to the size of the inhibitor can lead to complex sorbate–crystal binding (including the possibility of cooperative effects among multiple HCA molecules at step sites).

Inhibitor complexation of free calcium ions in solution reduces the rate of COM growth by lowering the supersaturation. The solubility of COM crystals C_s (that is, equilibrium concentration of solute) is as follows:

$$C_s = \sqrt{K_{sp} \gamma_{Ca}^{-1} \gamma_{Ox}^{-1}} \quad (11)$$

where K_{sp} is the solubility product ($1.66 \times 10^{-9}\,mol^2\,l^{-2}$ at 25 °C)[35] and γ_i is the activity coefficient (i = Ca or Ox). For the equimolar growth solutions used in this study ($C_{Ca} = C_{Ox}$), the activity coefficients for Ca and Ox are calculated using the expression:

$$\gamma = \exp\left(-\left(\frac{0.51 z^2 \sqrt{I}}{1 + \sqrt{I}} - 0.3I\right)\right) \quad (12)$$

where z is ion valence and I is ionic strength. The addition of either HCA or CA to a calcium oxalate growth solution lowers supersaturation by (i) forming complexes with free calcium ions to reduce C_{Ca}, and (ii) increasing the ionic strength, which alters the activity coefficients, thereby increasing the solubility of COM crystals. Inhibitor concentrations used in AFM experiments (about 3 μM) are insufficient to reduce supersaturation by complexation (that is, $Ca^{2+}/HCA \approx 10^3$), and impose only minor changes in solubility due to activity effects (that is, C_s increases by about 0.05%). If the inhibitor concentration in bulk crystallization is increased to values that are commensurate with supersaturation, the effects of complexation and activity are important. For example, we performed bulk crystallization in the presence of about 0.3 mM HCA (or 100 μg ml^{-1}) and observed few COM crystals, indicating almost complete suppression of COM crystallization.

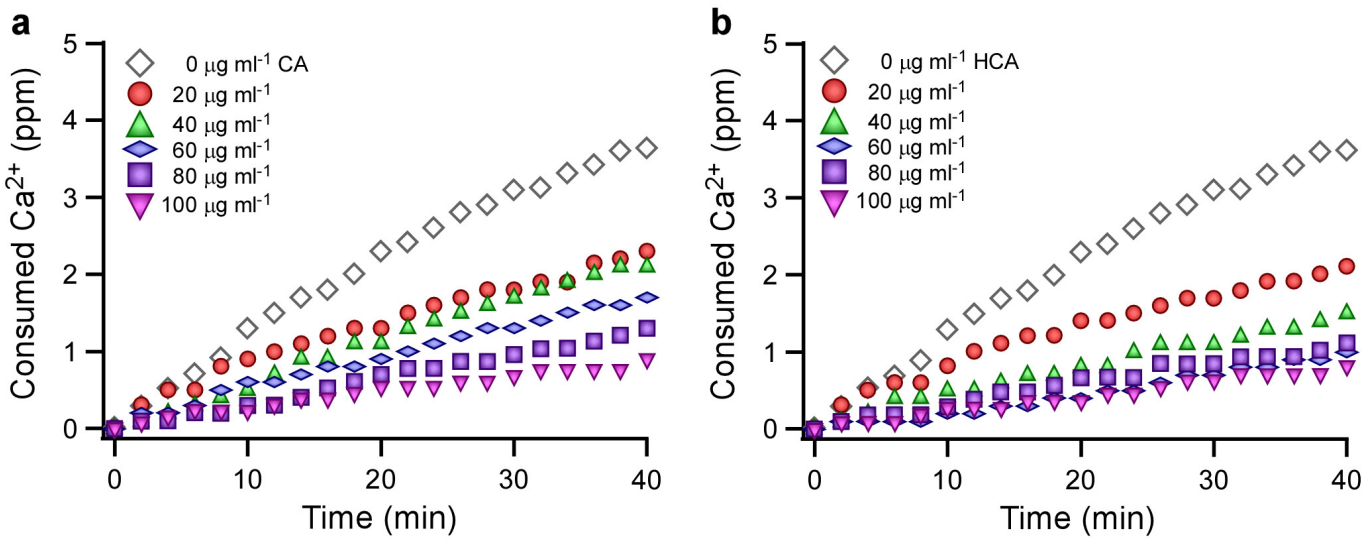
In the literature, alternative mechanisms have been proposed to describe the potential effects of inhibitors (or impurities) on crystal solubility. At low supersaturation, it is suggested that inhibitors can increase local solubility at crystal interfaces, potentially inducing dissolution when the mother liquor is close to saturation[61]. Under similar conditions, it is also possible to observe the effects of Ostwald ripening, where small crystals dissolve at the expense of larger particles that grow based on the Gibbs–Thompson effect[62]. Others have proposed that changes in surface free energy can be caused by the site occupancy in crystal lattices with embedded solid particles or impurities[63], which induce stress on the crystal (analogous to the simulations by ref. 26 described in Fig. 3a). Additionally, the formation of defects on crystal surfaces (such as vacancies or dislocations) can induce stress on the crystal lattice, leading to reduced rates of growth under supersaturated conditions[64].

34. Farmanesh, S. *et al.* High-throughput platform for design and screening of peptides as inhibitors of calcium oxalate monohydrate crystallization. *J. Cryst. Growth* **373,** 13–19 (2013).
35. Tomazic, B. B. & Nancollas, G. H. A study of the phase-transformation of calcium oxalate trihydrate-monohydrate. *Invest. Urol.* **16,** 329–335 (1979).
36. Lupulescu, A. I. & Rimer, J. D. In situ imaging of silicalite-1 surface growth reveals the mechanism of crystallization. *Science* **344,** 729–732 (2014).
37. Farmanesh, S. *et al.* Natural promoters of calcium oxalate monohydrate crystallization. *J. Am. Chem. Soc.* **136,** 12648–12657 (2014).
38. Jung, T. *et al.* Probing crystallization of calcium oxalate monohydrate and the role of macromolecule additives with in situ atomic force microscopy. *Langmuir* **20,** 8587–8596 (2004).
39. Ahlrichs, R., Bar, M., Haser, M., Horn, H. & Kolmel, C. Electronic-structure calculations on workstation computers—the program system turbomole. *Chem. Phys. Lett.* **162,** 165–169 (1989).
40. Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **38,** 3098–3100 (1988).
41. Perdew, J. P. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B* **33,** 8822–8824 (1986).
42. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **132,** 154104 (2010).
43. Mpourmpakis, G., Caratzoulas, S. & Vlachos, D. G. What controls Au nanoparticle dispersity during growth? *Nano Lett.* **10,** 3408–3413 (2010).
44. Feyereisen, M., Fitzgerald, G. & Komornicki, A. Use of approximate integrals in ab initio theory- An application in MP2 enegry calculations. *Chem. Phys. Lett.* **208,** 359–363 (1993).
45. Weigend, F. & Haser, M. RI-MP2: first derivatives and global consistency. *Theor. Chem. Acc.* **97,** 331–340 (1997).
46. Weigend, F., Haser, M., Patzelt, H. & Ahlrichs, R. RI-MP2: optimized auxiliary basis sets and demonstration of efficiency. *Chem. Phys. Lett.* **294,** 143–152 (1998).
47. Klamt, A. & Schuurmann, G. COSMO—a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc., Perkin Trans.* **2,** 799–805 (1993).
48. Millan, A. Crystal growth shape of whewellite polymorphs: influence of structure distortions on crystal shape. *Cryst. Growth Des.* **1,** 245–254 (2001).
49. Boll, M., Sorensen, E. & Balieu, E. Naturally occurring lactones and lactames. 3. Absolute configuration of hydroxycitric acid lactones—hibiscus acid and garcinia acid. *Acta Chem. Scand.* **23,** 286 (1969).
50. Asplin, J. R. Evaluation of the kidney stone patient. *Seminars Nephrol.* **28,** 99–110 (2008).
51. Farmanesh, S., Alamani, B. G. & Rimer, J. D. Identifying alkali metal inhibitors of crystal growth: a selection criterion based on ion pair hydration energy. *Chem. Commun.* **51,** 13964–13967 (2015).
52. Martell, A. E. & Smith, R. M. in *Critical Stability Constants* Vol. 5, Ch. XI, 329–330 (Plenum Press, 1982).
53. ChemAxon. *Hydroxycitric acid*. http://www.chemicalize.org/structure/#!mol=hydroxycitrate&source=fp (2016).
54. Cabrera, N. & Vermileya, D. A. in *Growth of Crystals from Solution* (eds Doremus R. H. *et al.*) Ch. V, 393–410 (Wiley, 1958).
55. Voronkov, V. V. & Rashkovich, L. N. Step kinetics in the presence of mobile adsorbed impurity. *J. Cryst. Growth* **144,** 107–115 (1994).
56. Burton, W. K., Cabrera, N. & Frank, F. C. The growth of crystals and the equilibrium structure of their surfaces. *Phil. Trans. R. Soc. Lond.* **243,** 299–360 (1951).
57. Frank, F. C. The influence of dislocations on crystal growth. *Discuss. Faraday Soc.* **5,** 48–54 (1949).
58. Lovette, M. A. *et al.* Crystal shape engineering. *Ind. Eng. Chem. Res.* **47,** 9812–9833 (2008).
59. Snyder, R. C. & Doherty, M. F. Predicting crystal growth by spiral motion. *Proc. R. Soc. A* **465,** 1145–1171 (2009).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

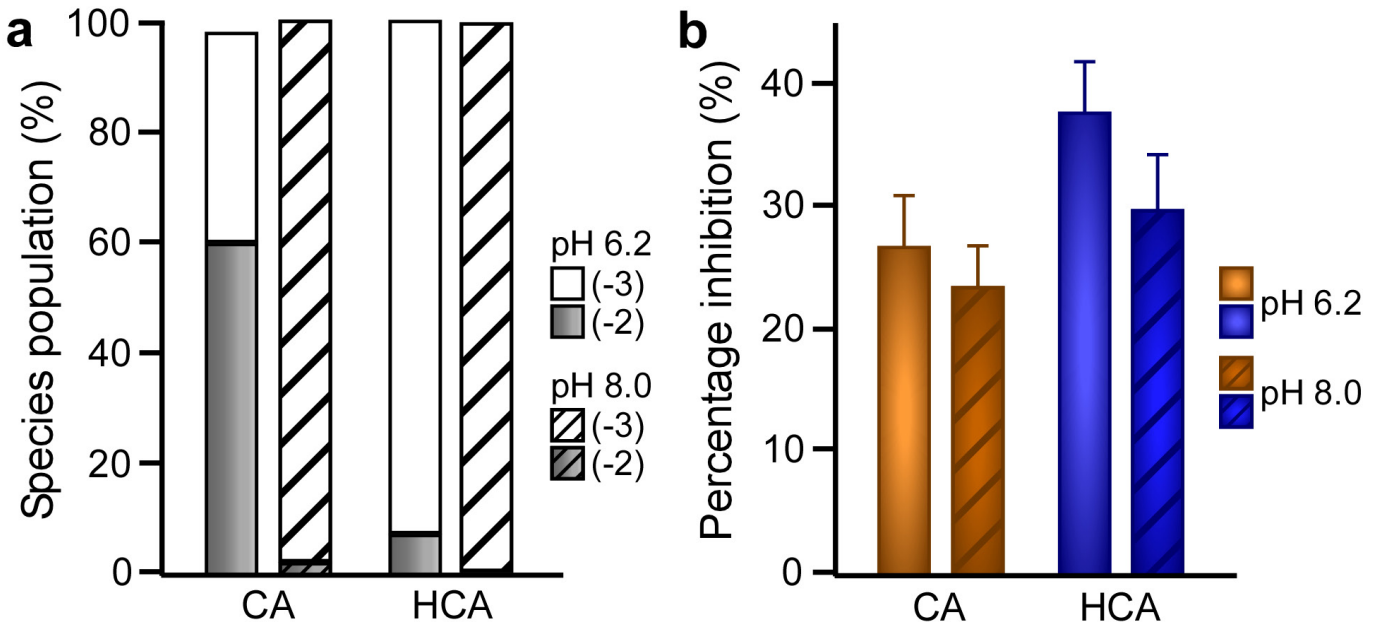
60. Winn, D. & Doherty, M. F. Modeling crystal shapes of organic materials grown from solution. *Am. Inst. Chem. Eng. J.* **46,** 1348–1367 (2000).
61. Linnikov, O. D. Investigation of the initial period of sulphate scale formation—Part 3. Variations of calcium sulphate crystal growth rates at its crystallization on a heat-exchange surface. *Desalination* **128,** 47–55 (2000).
62. Malivuk, D. A., Zekic, A. A., Mitrovic, M. M. & Misailovic, B. M. Dissolution of sodium chlorate crystals in supersaturated solutions. *J. Cryst. Growth* **377,** 164–169 (2013).
63. Cahn, J. W. & Larche, F. Surface stress and the chemical equilibrium of small crystals—II. Solid particles embedded in a solid matrix. *Acta Metall.* **30,** 51–56 (1982).
64. van der Heijden, A. E. D. M. & van der Eerden, J. P. Growth rate dispersion: the role of lattice strain. *J. Cryst. Growth* **118,** 14–26 (1992).
65. Wang, L. J. *et al.* Constant composition studies verify the utility of the cabrera-vermilyea (C-V) model in explaining mechanisms of calcium oxalate monohydrate crystallization. *Cryst. Growth Des.* **6,** 1769–1775 (2006).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



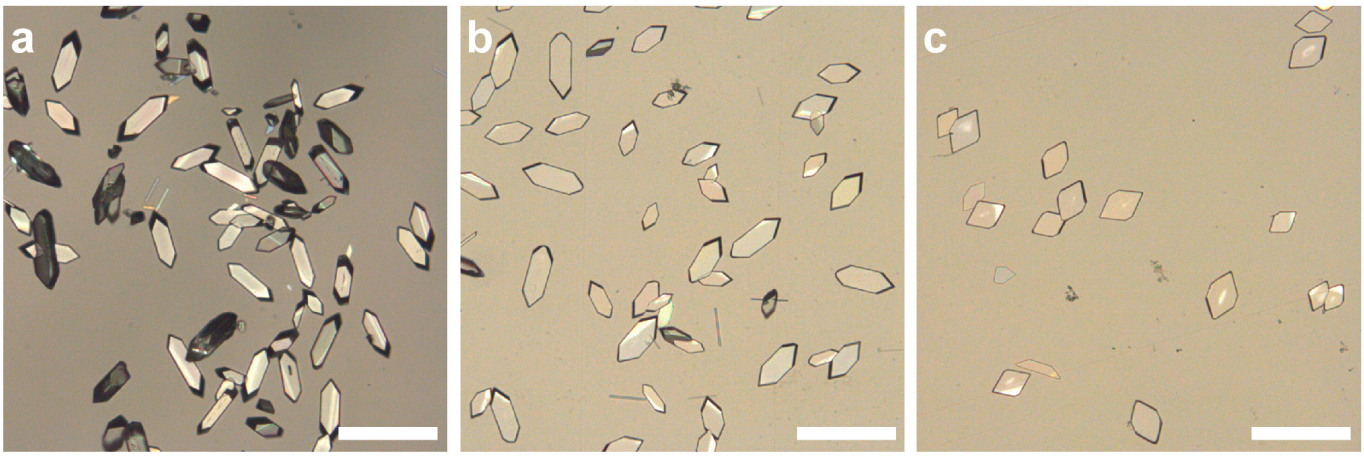
Extended Data Figure 1 | Examples of ISE measurements. *In situ* ISE measurements of COM crystallization in the presence of CA (**a**) and HCA (**b**) at concentrations of 0 μg ml^{-1}, 20 μg ml^{-1}, 40 μg ml^{-1}, 60 μg ml^{-1}, 80 μg ml^{-1} and 100 μg ml^{-1}. The *y* axis is the quantity of free calcium ions in the growth solution that are consumed during crystallization. Linear regression of each curve provides the rate of crystal growth. The percentage inhibition of COM crystallization is obtained by comparing the slopes of ISE curves in the presence of inhibitor (filled symbols) to the slope in the absence of inhibitor (open diamonds). ppm, parts per million.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



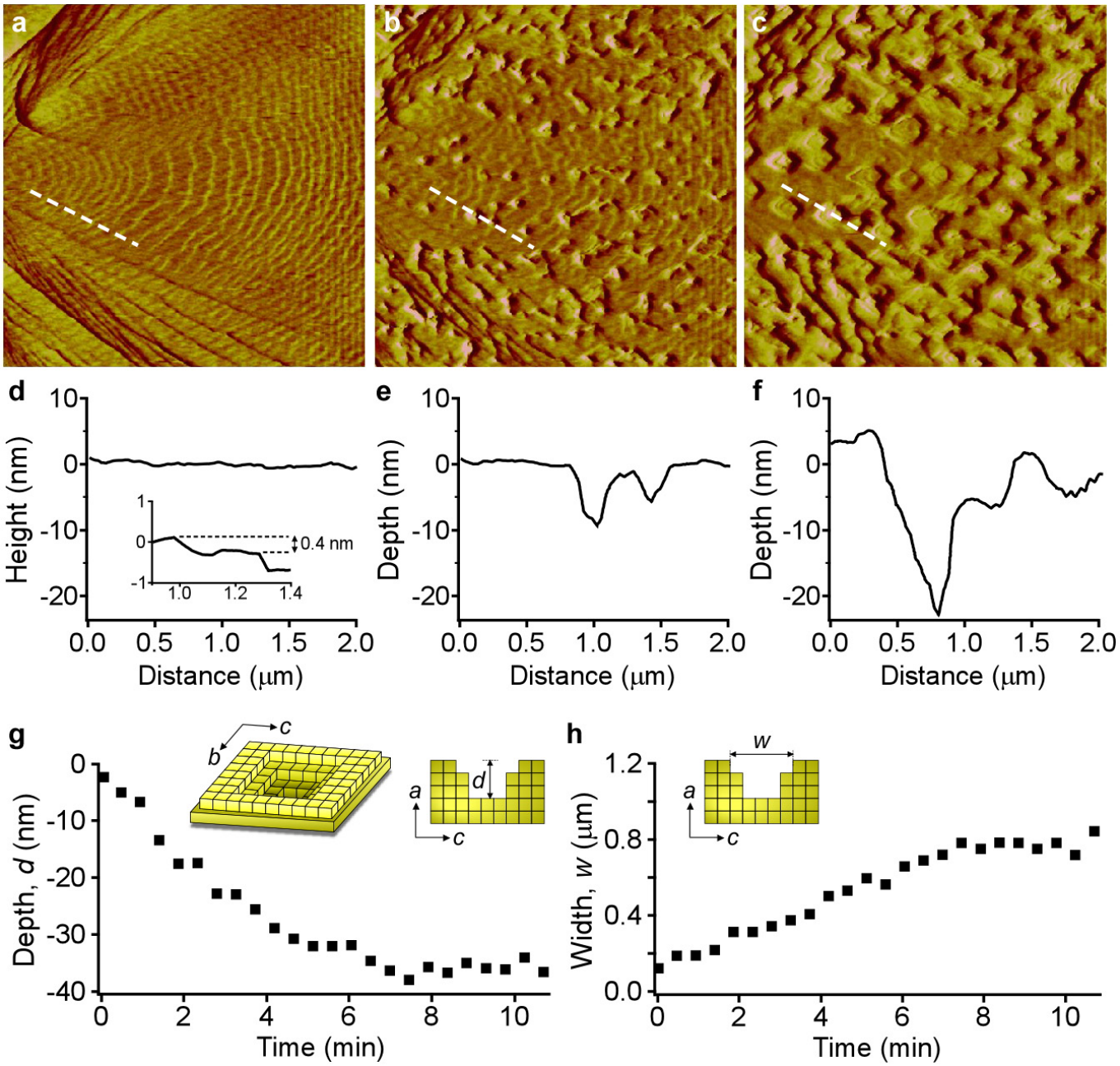
Extended Data Figure 2 | Inhibitor speciation and its effect on COM growth. Here we compare calculations of inhibitor speciation with ISE measurements of COM percentage inhibition at pH 6.2 (solid bars) and pH 8.0 (patterned bars). **a**, Percentage of deprotonated CA and HCA species, calculated from equations (4)–(9). Fully dissociated species (charge −3) are represented by white bars and partially dissociated species (charge −2) are represented by grey bars. **b**, Results of ISE measurements in supersaturated calcium oxalate solution ($S = 3.8$) in the presence of $C_{CA} = 20\,\mu g\,ml^{-1}$ (orange bars) and $C_{HCA} = 20\,\mu g\,ml^{-1}$ (blue bars). The percentage inhibition of COM crystal growth slightly decreases at higher alkalinity. HCA is the more effective inhibitor irrespective of solution pH. Data are the average of more than 10 measurements (error bars are 1 s.d.; $P < 0.05$ comparing HCA to CA at both levels of pH).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



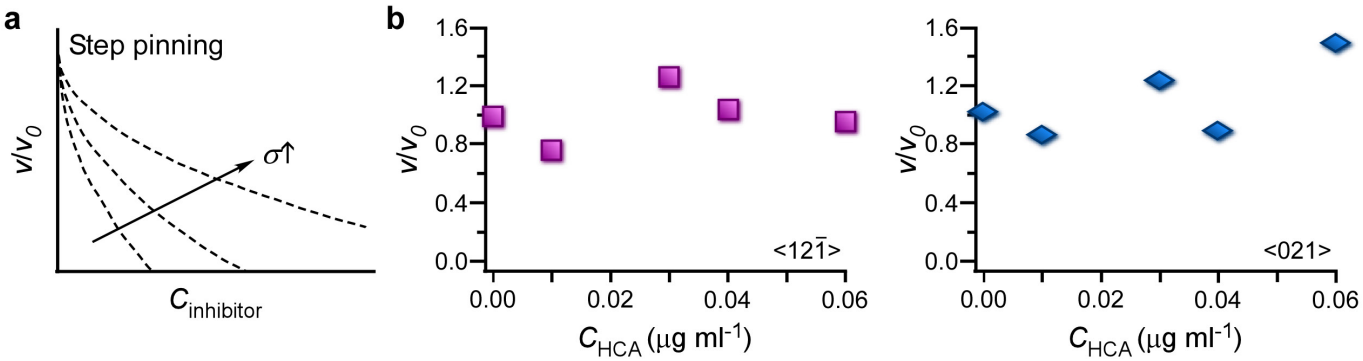
Extended Data Figure 3 | Optical micrographs of COM crystals. Optical micrographs of COM crystals after heating a growth solution for 3 days at 60 °C. Here we compare the control sample in the absence of growth inhibitor (**a**) to solutions prepared with $C_{CA} = 20\,\mu g\ ml^{-1}$ (**b**) and $C_{HCA} = 20\,\mu g\ ml^{-1}$ (**c**). Scale bars, 100 μm.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



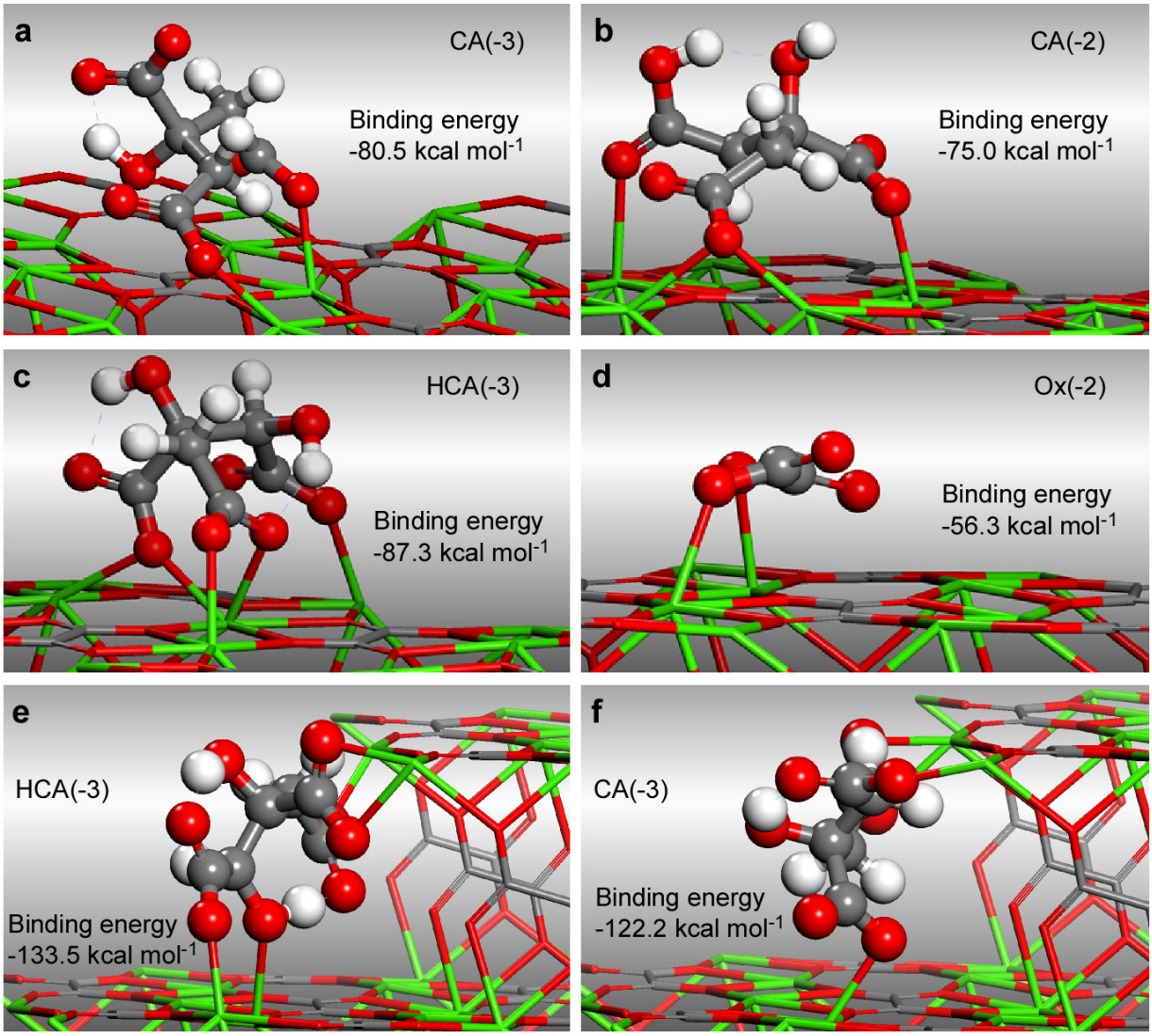
Extended Data Figure 4 | HCA-induced etch pits on the COM (100) surface. a–c, Time-resolved images during *in situ* AFM measurements of a COM (100) crystal surface in the presence of HCA. The surface is first imaged in the absence of inhibitor (**a**) and then at $C_{HCA} = 0.25\,\mu g\ ml^{-1}$ (**b** and **c**). The elapsed time between each deflection mode image is approximately 4 min. **d–f**, Height (or depth) profiles corresponding to the dashed lines in **a–c**, respectively. As shown in **a** and **d**, the COM (100) surface before the addition of HCA is comprised of single steps with height approximately 0.4 nm (see inset in **d**), which approximately corresponds to the unit cell dimension in the [100] direction ($a = 0.6$ nm)[48]. Depth profiles in **e** and **f** show the temporal evolution of a single etch pit. Quantitative analysis of etch pit dimensions with respect to depth d (**g**) and width w (**h**) reveal monotonic changes during 10 min of continuous AFM imaging. Schematics of an etch pit (left inset in **g**) with highlighted depth (right inset in **g**) and height (inset in **h**) are shown to aid visualization.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



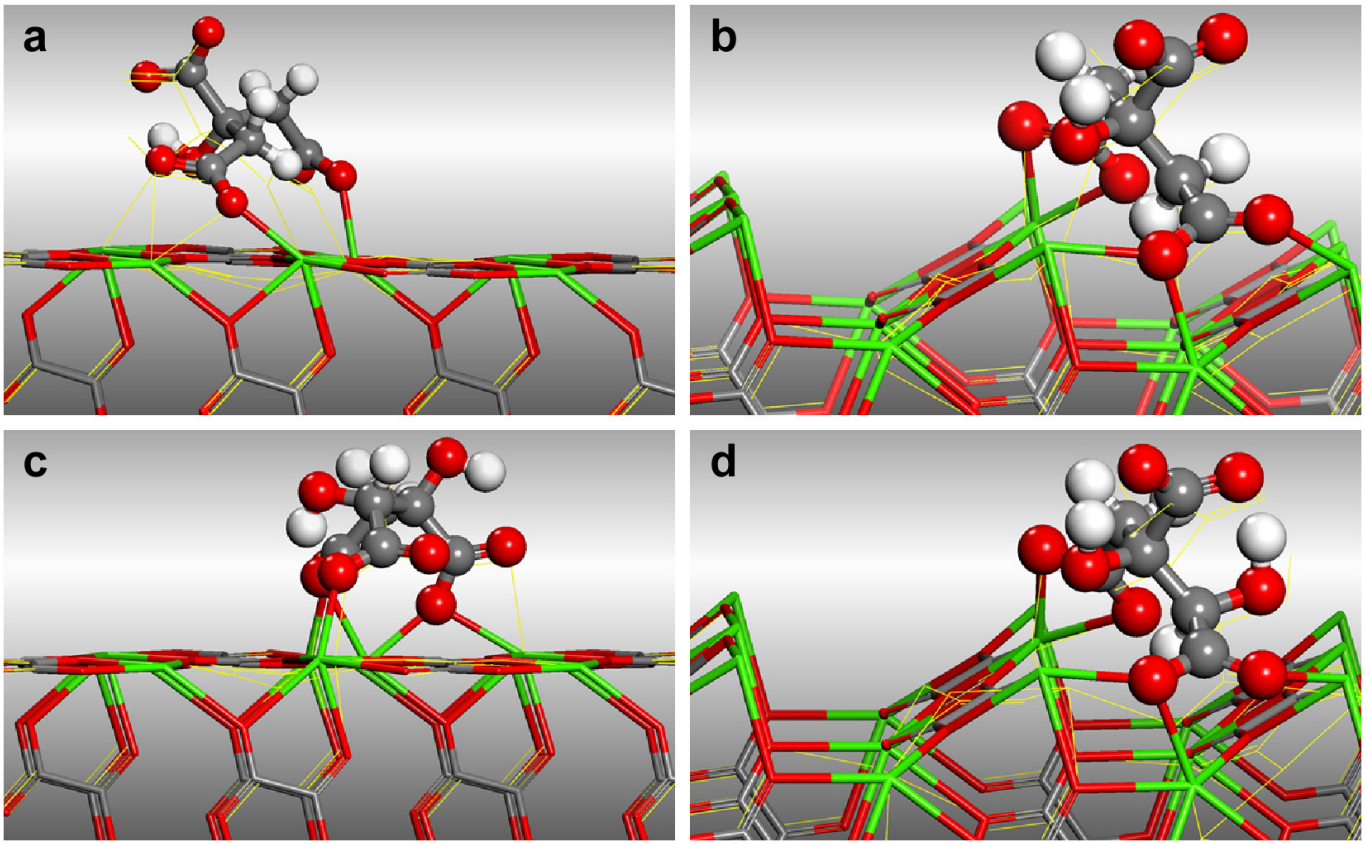
Extended Data Figure 5 | Construction of Bliznakov plots. a, Theoretical Bliznakov plot for crystal inhibitors that follow a step-pinning mode of action as a function of increasing calcium oxalate relative supersaturation σ (derived from ref. 65). **b**, Plots generated from *in situ* AFM data on COM (010) surfaces compare changes in step velocity in the [12$\bar{1}$] direction (purple squares) and [021] direction (blue diamonds) as a function of increasing C_{HCA}. The deviation of experimental data from theoretical trends suggests that step pinning is not the dominant mechanism by which HCA inhibits COM surface growth.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



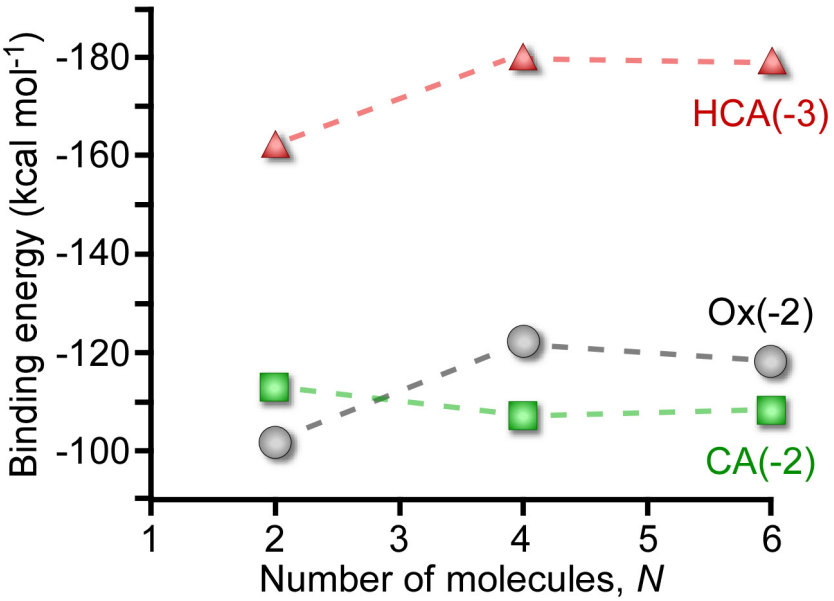
Extended Data Figure 6 | Binding energy of inhibitors on the COM (100) surface. The results of DFT calculations showing the adsorption configuration and binding energy of CA^{3-} (**a**), CA^{2-} (**b**), HCA^{3-} (**c**) and Ox^{2-} (**d**) on the COM (100) surface and of HCA^{3-} (**e**) and CA^{3-} (**f**) binding to a (001) step on the COM (100) surface. For these calculations, the surfaces are kept frozen (that is, unrelaxed). Atoms are coloured as follows: hydrogen (white), carbon (grey), oxygen (red) and calcium (green).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 7 | CA and HCA interactions with relaxed COM surfaces. Superimposed structures of CA and HCA interacting with unrelaxed (coloured balls and sticks) and partially relaxed (yellow sticks) surfaces of COM crystals. Side-view snapshots depict CA interaction with (100) (**a**) and (021) (**b**) surfaces and HCA interaction with (100) (**c**) and (021) (**d**) surfaces. Atoms are coloured as follows: hydrogen (white), carbon (grey), oxygen (red) and calcium (green). The (100) surface is practically unaffected by the presence of HCA, whereas the (021) surface shows dislocations due to strain induced by the high binding affinity of the inhibitor (see also Extended Data Table 1). The total energy of the (100) face changes by $+18.8\,\text{kcal mol}^{-1}$ owing to HCA adsorption compared to the $+28.1\,\text{kcal mol}^{-1}$ energy change of the (021) face (positive signs are endothermic and energy values correspond to the difference between single point energy calculations of the COM surface with inhibitors removed). The corresponding values for the total energy change of the (100) and (021) faces from the presence of the CA are $+14.2\,\text{kcal mol}^{-1}$ and $+25.7\,\text{kcal mol}^{-1}$, respectively. The partial relaxation of COM surfaces compared to unrelaxed surface calculations (Extended Data Fig. 6) does not alter the overall trend in inhibitor–crystal binding affinity.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 8 | Complexation of organic acids with calcium. DFT-calculated binding energy (scaled per number of molecules, *N*) for the complexation of organic anions HCA^{3-}, CA^{2-} and Ox^{2-} with calcium ions. Note that the data for HCA^{3-} and Ox^{2-} are identical to Fig. 3g, and are merely placed here for direct comparison with CA^{2-}. Dashed lines connecting symbols are added to guide the eye.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 1 | Quantification of lattice strain in the presence of HCA and CA

	Average displacement on COM (100), δ (values in Å)	Average displacement on COM (021), δ (values in Å)
Pristine	0.17	0.38
CA(-3)	0.25	0.55
HCA(-3)	0.25	0.71

Average displacement of (100) and (021) surface sites during partial surface relaxation where the pristine surfaces (absence of modifier) are compared to surfaces with one adsorbed molecule of HCA or CA, calculated relative to the frozen surface.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

LETTER

doi:10.1038/nature19100

An integrated design and fabrication strategy for entirely soft, autonomous robots

Michael Wehner[1,2]*, Ryan L. Truby[1,2]*, Daniel J. Fitzgerald[1,2], Bobak Mosadegh[3,4], George M. Whitesides[2,5], Jennifer A. Lewis[1,2] & Robert J. Wood[1,2]

Soft robots possess many attributes that are difficult, if not impossible, to achieve with conventional robots composed of rigid materials[1,2]. Yet, despite recent advances, soft robots must still be tethered to hard robotic control systems and power sources[3–10]. New strategies for creating completely soft robots, including soft analogues of these crucial components, are needed to realize their full potential. Here we report the untethered operation of a robot composed solely of soft materials. The robot is controlled with microfluidic logic[11] that autonomously regulates fluid flow and, hence, catalytic decomposition of an on-board monopropellant fuel supply. Gas generated from the fuel decomposition inflates fluidic networks downstream of the reaction sites, resulting in actuation[12]. The body and microfluidic logic of the robot are fabricated using moulding and soft lithography, respectively, and the pneumatic actuator networks, on-board fuel reservoirs and catalytic reaction chambers needed for movement are patterned within the body via a multi-material, embedded 3D printing technique[13,14]. The fluidic and elastomeric architectures required for function span several orders of magnitude from the microscale to the macroscale. Our integrated design and rapid fabrication approach enables the programmable assembly of multiple materials within this architecture, laying the foundation for completely soft, autonomous robots.

Soft robotics is a nascent field that aims to provide safer, more robust robots that interact with humans and adapt to natural environments better than do their rigid counterparts. Unlike conventional robots composed of rigid materials, soft robots based on hydrogels[15,16], electroactive polymers[17], granular media[18] and elastomers[5,19] exhibit elastic moduli ranging from 10 kPa to 1 GPa (ref. 1), are physically resiliant[7,20] and have the ability to passively adapt to their environment[1,2,19]. Moulded and laminated elastomers with embedded pneumatic networks are widely used materials in soft robotics[1,21,22]. Actuation of these elastomeric composites occurs when interconnected channels that make up the pneumatic network are inflated with incompressible fluids or gases supplied via tethered pressure sources[1]. Robotic end effectors with bioinspired[10] and rapid[6] actuation, deployable crawlers[3,7] and swimmers[8] with complex body motions, and robust jumpers[9,23] have been developed on the basis of this design strategy. However, in each case, these robots are either tethered to or carry rigid systems for power and control, yielding hybrid soft–rigid systems[4,7–9].

Creating a new class of fully soft, autonomous robots[24] is a grand challenge, because it requires soft analogues of the control and power hardware currently used. Recently, monopropellant fuels have been suggested as a promising fuel source for pneumatically actuated soft robots[4,12]. Their rapid decomposition into gas upon exposure to a catalyst offers a strategy for powering soft robotic systems that obviates the need for batteries or external power sources. Here, we report a method for creating a completely soft, pneumatic robot—the 'octobot'—with eight arms that are powered by monopropellant decomposition.

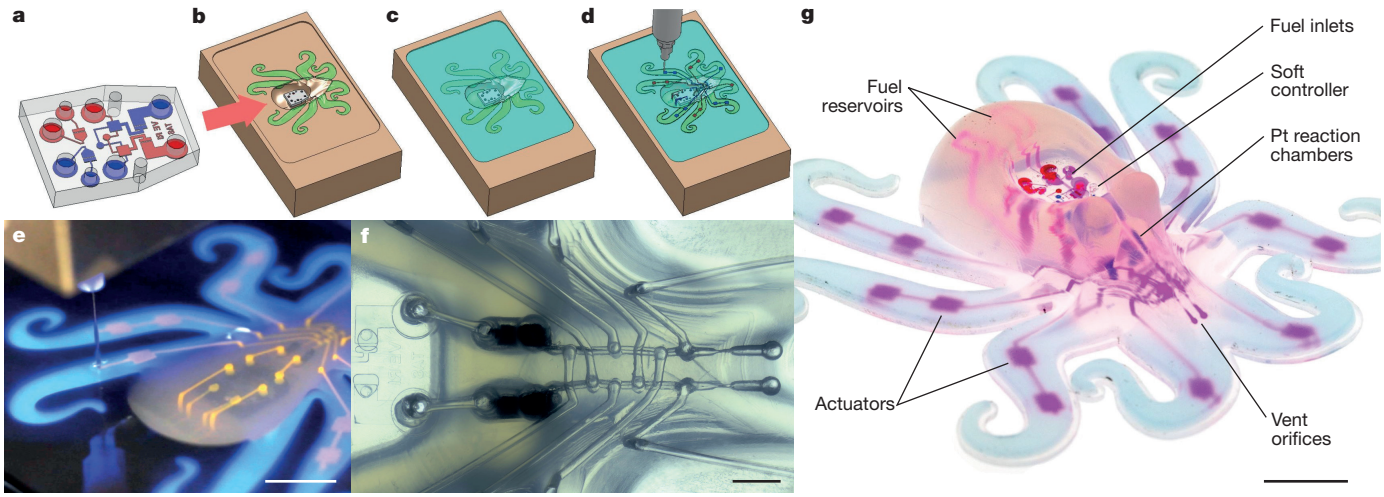


Figure 1 | Fully soft, autonomous robot assembly. a, **b**, A microfluidic soft controller is pre-fabricated (**a**) and loaded into a mould (**b**). **c**–**e**, Matrix materials are poured into the mould (**c**) and fugitive and catalytic inks are EMB3D printed (**d**, **e**). Scale bar in **e**, 10 mm. **f**, After matrix curing, the printed fugitive inks auto-evacuate, yielding open channels. Scale bar, 2 mm. **g**, The octobot is removed from the mould and inverted to reveal a fully soft, autonomous robot that is controlled via the embedded microfluidic soft controller and powered by monopropellant decomposition. Scale bar, 10 mm. Fluorescent dyes have been added in **e** and **g** to assist in visualization of internal features.

[1]John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA. [2]Wyss Institute for Biologically Inspired Engineering, Harvard University, Cambridge, Massachusetts 02138, USA. [3]Dalio Institute of Cardiovascular Imaging, Weill Cornell Medicine and New York Presbyterian Hospital, New York, New York 10021, USA. [4]Department of Radiology, Weill Cornell Medicine, New York, New York 10021, USA. [5]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, USA.
*These authors contributed equally to this work.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

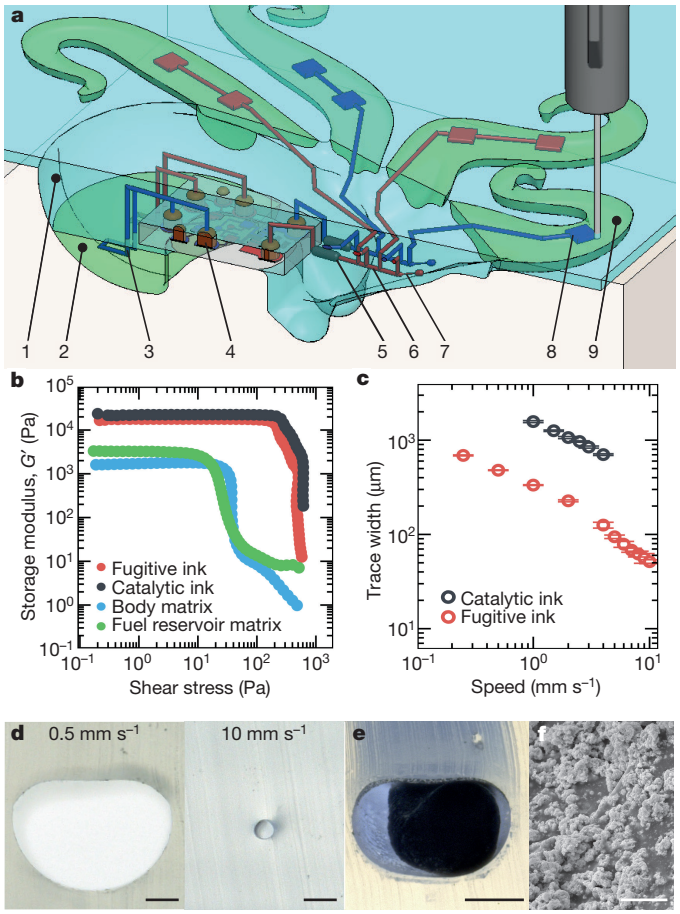


Figure 2 | Multi-material, EMB3D printing. a, The octobot features include (1) the body matrix, (2) the fuel reservoir matrix, (3) printed fuel reservoir traces, (4) fugitive plugs in the soft controller, (5) printed platinum reaction chambers, (6) printed pneumatic networks, (7) printed vent orifices, (8) printed actuators and (9) moulded hyperelastic actuator matrix. All printed features are composed of the fugitive ink except the printed platinum reaction chambers (5), which are patterned using the catalytic ink. **b**, The storage modulus, G', of the fugitive ink, catalytic ink, body matrix and fuel reservoir matrix as a function of shear stress. The plateau moduli of the inks are an order of magnitude higher than those of the matrix materials. **c**, Trace widths of the fugitive and catalytic inks printed at 450 kPa and 345 kPa, respectively, decrease with print speed. Error bars indicate the standard deviation for $n = 3$ measurements. **d**, Optical images of channel cross-sections printed at speeds of 0.5 mm s^{-1} and 10 mm s^{-1}, which demonstrate that trace dimensions can be changed on-the-fly. Scale bars, 100 μm. **e**, **f**, Reaction chambers printed with the catalytic inks contain a platinum-laden plug, as shown in a cross-section (**e**; scale bar, 500 μm) and a scanning electron micrograph (**f**; scale bar, 25 μm).

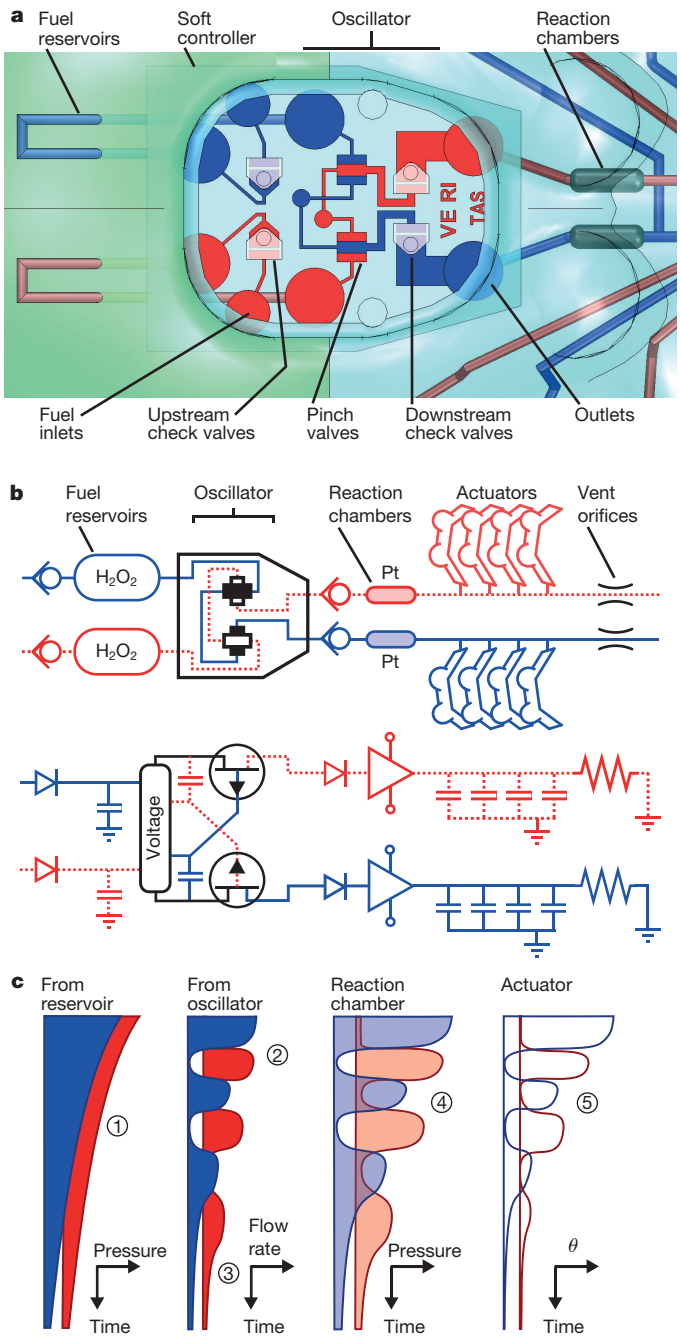


Figure 3 | Octobot control logic. Discrete sides are shown in red and blue for clarity. **a**, A system of check valves and switch valves within the soft controller regulates fluid flow into and through the system. The letters of 'VERITAS', each with a height of 500 μm, are patterned into the soft controller as an indication of scale. **b**, A schematic (top) and qualitative electrical analogy (bottom) of the octobot system; check valves, fuel tanks, oscillator, reaction chambers, actuators and vent orifices are akin to diodes, supply capacitors, electrical oscillator, amplifiers, capacitors and pull-down resistors, respectively. **c**, Conceptual curves show key variables as a function of time. (1) Nominal pressure drives fuel through system at a decreasing rate. (2) Pinch valves in the oscillator convert upstream flow into alternating flow between red and blue channels. Flow rate and switching frequency are functions of upstream pressure and downstream impedance. (3) When upstream pressure is too low, oscillation is not possible, so both sides flow at a reduced rate. (4) Catalyst decomposes fuel, yielding pressurized gas, which flows downstream to the actuators and the vent orifices concurrently. (5) Actuators deform (θ, actuator tip angle) as the pressure changes. Vents must be sufficiently small to allow full actuation, yet sufficiently large to allow timely venting.

To accomplish this, we use microfluidic logic[11] as a soft controller and a multi-material, embedded 3D (EMB3D) printing method to fabricate pneumatic networks within a moulded, elastomeric robot body. Our hybrid assembly approach allows one to seamlessly integrate soft lithography, moulding and 3D printing to rapidly and programmably fabricate a range of materials and functional elements in the form factors that are required for autonomous, untethered operation of a soft robot.

To fabricate an octobot, we first micro-mould[11,25] the soft controller that houses the microfluidic logic necessary for controlling fuel decomposition (Fig. 1a). The soft controller, which is protected temporarily with a polyimide mask, is placed into a mould that is partially filled by hyperelastic layers, which are needed for actuation (Fig. 1b). Matrix materials are then poured into the mould (Fig. 1c) and the remaining soft robot features are EMB3D printed into the moulded matrix (Fig. 1d, e, Supplementary Video 1). After the matrix materials are cross-linked, the aqueous inks 'auto-evacuate' at elevated temperature

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

as water evaporates and diffuses through the matrix, leaving behind an open network of channels that are interfaced with the soft controller (Fig. 1f). Octobot fabrication is completed upon removal of excess matrix material (Fig. 1g). A more detailed description of this multi-step assembly process is provided in Extended Data Fig. 1.

By combining micro-moulding with EMB3D printing, we rapidly patterned the required mesofluidic networks by extruding a fugitive ink, which is subsequently removed via auto-evacuation, through a fine nozzle that is embedded within the uncured elastomer matrix. To self-heal crevices that form within the 'body' matrix as the nozzle is translated during the printing process, we created a new elastomeric material containing fumed silica nanoparticles that exhibits thixotropic behaviour[26] (Extended Data Fig. 2a). When completely restructured or at rest, this matrix behaves like a Herschel–Bulkley fluid; that is, it exhibits both shear-thinning behaviour (Extended Data Fig. 2b) and a shear yield stress (Extended Data Fig. 2c). These properties ensure that the extruded inks remain in place within the matrix[13,14]. However, upon yielding, the body matrix readily flows (Extended Data Fig. 2c) into any crevices formed. The body matrix restructures with time, ultimately recovering its original viscoelasticity (Extended Data Fig. 3), which ensures that EMB3D printing can be repeated later in the same matrix region. We also created a 'fuel reservoir' elastomeric matrix, into which fuel reservoir channels are printed. Both the body and fuel reservoir matrices are cross-linked within the mould after printing is completed.

To create the fuel reservoirs, catalytic reaction chambers, actuator networks and vent orifices, two hydrogel-based inks (fugitive and catalytic) are EMB3D printed within the moulded matrix materials (Fig. 2a). These printed features are interfaced with each other as well as with the soft controller through the use of 'fugitive plugs' introduced at the inlets of the controller before filling the mould with the matrix materials. The fugitive ink is composed of an aqueous, poly(ethylene oxide)-*b*-poly(propylene oxide)-*b*-poly(ethylene oxide) triblock copolymer (Pluronic F127) gel[13,27]. The catalytic ink contains platinum particles (Supplementary Video 2) suspended in a mixture of Pluronic F127-diacrylate (F127-DA) and poly(ethylene glycol) diacrylate (PEG-DA) that is photo-cross-linked after printing. The rheological properties of both inks are specifically tailored for EMB3D printing[13,14] (Fig. 2b, Extended Data Fig. 4). The printed features produced from both inks can be changed 'on-the-fly' by varying the print speed (Fig. 2c). Typically, this fugitive ink must be removed or evacuated after printing to yield open channels[13,27]. However, we find that fugitive ink composed of pure Pluronic F127 can be auto-evacuated by heating the printed features within the cross-linked, silicone-based matrices at 90 °C (ref. 28; Fig. 2d, Extended Data Fig. 5). As water evaporation

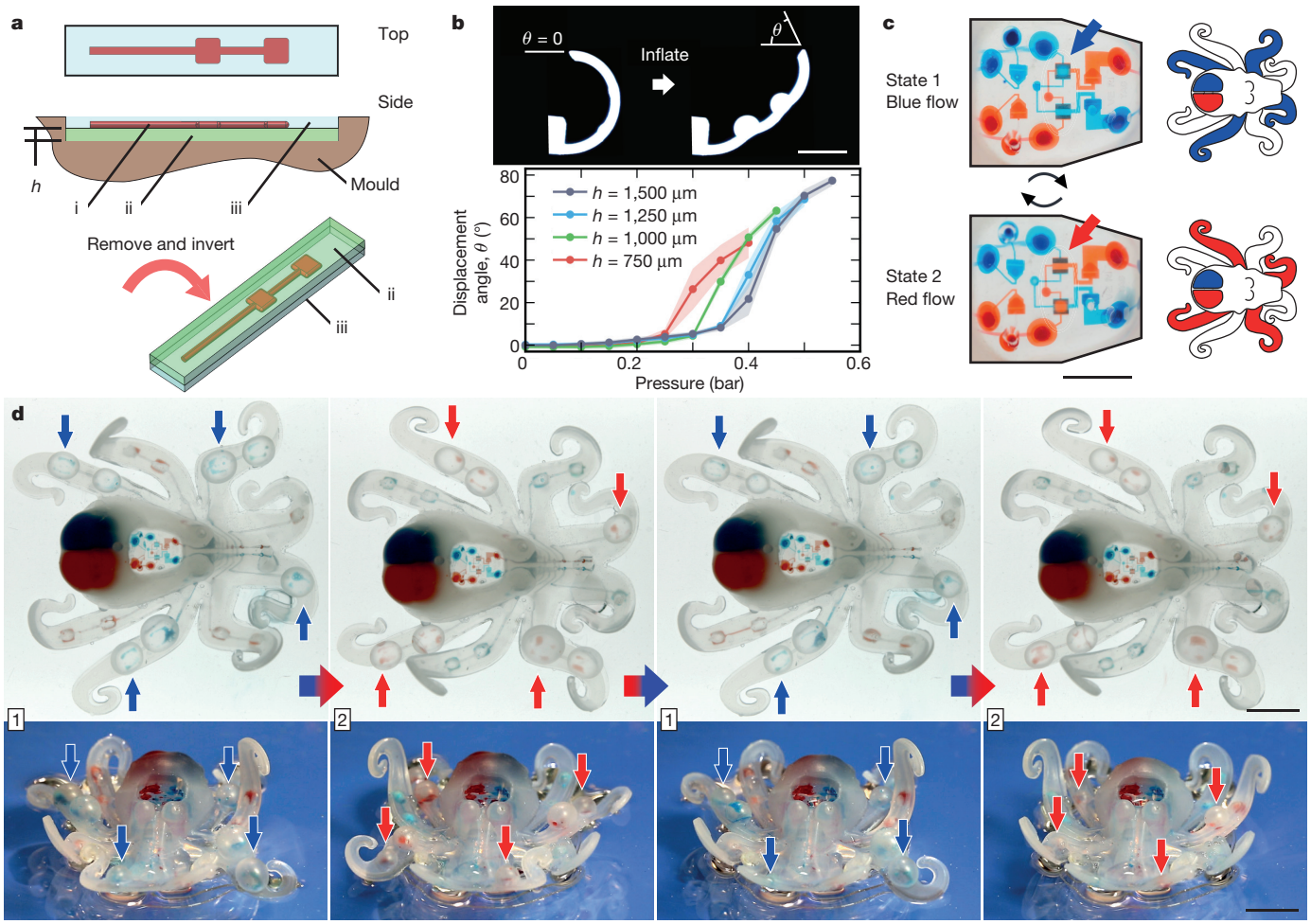


Figure 4 | Octobot actuation. a, Two-bladder actuator design in which traces (i) are printed in contact with the hyperelastic layer (ii) inside the body matrix material (iii) and differences in modulus result in bending upon inflation. The thickness, *h*, of the hyperelastic layer is modified to change the characteristics of the actuator. In this example, the body matrix material (iii) has a height of 800 μm. **b**, Top, the actuator tip angle, *θ*, changes upon inflation. Scale bar, 10 mm. Bottom, mean displacement angle, *θ*, taken from three representative actuators during five inflation cycles as a function of inflation pressure, for varying hyperelastic layer heights, *h* (in μm). Error bars, denoted by the shaded regions, indicate the 95% confidence interval. **c**, The oscillator of the soft controller causes an octobot to alternate between blue and red actuation states. The monopropellant fuel is dyed to show states. Scale bar, 5 mm. **d**, Stills from top-down (top; Supplementary Video 5) and face-on (bottom; Supplementary Video 6) operation videos show an octobot autonomously alternating between blue ('1') and red ('2') actuation states. Scale bars, 10 mm.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

ensues, the triblock copolymer species either form a thin coating at the matrix-open channel interface or may partially diffuse into the matrix. The fugitive plugs within the inlets of the soft controller also undergo this auto-evacuation process, facilitating connectivity between the microfluidic logic and all printed mesofluidic components (Extended Data Fig. 6). By contrast, the catalytic ink is cross-linked in place after printing, yielding a platinum-laden plug within the matrix (Fig. 2e).

To achieve the desired autonomous function, we incorporated a soft, microfluidic controller within the octobot (Fig. 3a). The control system is roughly divided into four sections: upstream (liquid fuel storage), oscillator (liquid fuel regulation), reaction chamber (decomposition into pressurized gas) and downstream (gas distribution for actuation and venting). Upstream, 0.5 ml of fuel is infused via a syringe pump into each of two fuel reservoirs printed into the hyperelastic matrix. Upstream check valves in the soft controller prevent fuel from flowing back out the fuel inlets. The fuel reservoirs expand elastically to a pressure of approximately 50 kPa, forcing fuel into the oscillator. The oscillator includes a system of pinch and check valves based on prior designs[11], which convert pressurized fuel inflow into alternating fuel outflow. With one channel temporarily occluded, fuel from the other channel flows from the outlets of the soft controller into the platinum-laden reaction chambers, where it rapidly decomposes. The resulting pressurized gas, which is prevented from returning to the soft controller via downstream check valves, flows into one of the downstream mesofluidic networks consisting of four actuators and one orifice. The supplied pressure deflects the actuators and exhausts to atmosphere through the vent orifice. Therefore, for robust actuation and timely venting, a balance must be reached between supply gas flow, actuation pressure and exhaust rate. These subcomponents operate on the basis of the interaction and timing of the local pressures, which is conceptually similar to an electrical oscillator (Fig. 3b). Upon successful venting, the fuel flow into one reaction chamber stops and flow to the other begins, initiating a similar sequence in the other downstream catalytic chamber and actuator network (Fig. 3c).

To provide an on-board power source, we use 50 wt% aqueous hydrogen peroxide as the fuel, owing to its high energy density (1.44 kJ g^{-1} as compared to 0.1–0.2 kJ g^{-1} for batteries) and its benign decomposition by-products. As the fuel decomposes in the presence of the platinum catalyst, the following reaction occurs: $2H_2O_2\,(l) \rightarrow 2H_2O\,(l, g) + O_2\,(g)$. This reaction results in volumetric expansion by a factor of approximately 240 (at ambient pressure)[12]. At our operating pressure of 50 kPa, an expansion to 160 times the original volume is expected. Although higher fuel concentrations would provide increased expansion and energy density, concentrations above 50 wt% drastically increase the decomposition temperature, resulting in combustion within the printed catalytic reaction chambers (Supplementary Videos 3 and 4). Because this monopropellant liquid fuel can be handled in small volumes and decomposes at the point of use, we can use microfluidic logic to directly handle the fuel, eliminating the need for external valves[7] to control gas at high pressure and flow rate.

The geometry of the microfluidic soft controller is designed to operate at a fuel flow rate of about 40 μl min^{-1}, thereby yielding pressurized gas at a rate of about 6.4 ml min^{-1} (ref. 11). Under these operating conditions, the theoretical run time of 12.5 min could be achieved using a system with a fuel capacity of 1 ml. The actuators, which consist of printed bladders in contact with a lower-modulus, hyperelastic elastomer layer (Fig. 4a), are designed to inflate asymmetrically to generate angular displacement. Their maximum working pressure and displacement are tuned on the basis of the thickness of the hyperelastic layer (Fig. 4b, Extended Data Fig. 7). If this layer is too thin, then it ruptures prematurely. However, the working pressure increases with thickness. As a compromise, we selected a layer thickness of 1,000 μm because it affords consistent performance at the lowest working pressure. In parallel with the actuators, we tailored the diameter of the vent orifices by modulating print speed. Orifices roughly 75 μm in width allowed proper actuator displacement with timely subsequent venting. The ability to rapidly pattern and adjust the geometry of these features on-the-fly via EMB3D printing allowed us to iterate through more than 30 designs and nearly 300 octobots to converge on an appropriate system-level architecture.

Through this iterative process, we created octobots with embedded components that work together in concert to alternate between the red and blue actuation states shown in Fig. 4c. The resulting octobots operated autonomously (Fig. 4d, Extended Data Fig. 8, Supplementary Videos 5 and 6), cycling between actuation states for four to eight minutes. Although this is less than the predicted theoretical run time, the soft controller alternates actuation states as expected. We believe that downstream impedances arising from decomposition–actuation–venting cycles, as well as the decreasing flow rate of fuel into the soft controller with time, are responsible for the departure from theoretical performance[11]. These issues can be addressed by integrating more sophisticated microfluidic circuits, such as universal logic gates, or components with 'gain' that enable advanced control schemes (see Methods for an extended discussion).

We have demonstrated the untethered operation of a robot composed solely of soft materials. The coupling of monopropellant fuels and microfluidic logic allowed us to power, control and realize autonomous operation of these pneumatically actuated systems. Through our hybrid assembly approach, we both constructed the robot body and embedded the necessary components for fuel storage, catalytic decomposition and actuation to enable system-level function in a rapid manner. The octobot is a minimal system designed to demonstrate our integrated design and fabrication strategy, which may serve as a foundation for a new generation of completely soft, autonomous robots.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 29 March; accepted 7 July 2016.

1. Rus, D. & Tolley, M. T. Design, fabrication and control of soft robots. *Nature* **521,** 467–475 (2015).
2. Wang, L. & Iida, F. Deformation in soft-matter robotics: a categorization and quantitative characterization. *IEEE Robot. Autom. Mag.* **22,** 125–139 (2015).
3. Shepherd, R. F. *et al.* Multigait soft robot. *Proc. Natl Acad. Sci. USA* **108,** 20400–20403 (2011).
4. Onal, C. D., Chen, X., Whitesides, G. M. & Rus, D. Soft mobile robots with on-board chemical pressure generation. *In 15th International Symposium on Robotics Research (ISRR 2011)* 1–16 (2011).
5. Lin, H.-T., Leisk, G. G. & Trimmer, B. GoQBot: a caterpillar-inspired soft-bodied rolling robot. *Bioinspir. Biomim.* **6,** 026007 (2011).
6. Mosadegh, B. *et al.* Pneumatic networks for soft robotics that actuate rapidly. *Adv. Funct. Mater.* **24,** 2163–2170 (2014).
7. Tolley, M. T. *et al.* A resilient, untethered soft robot. *Soft Robot.* **1,** 213–223 (2014).
8. Marchese, A. D., Onal, C. D. & Rus, D. Autonomous soft robotic fish capable of escape maneuvers using fluidic elastomer actuators. *Soft Robot.* **1,** 75–87 (2014).
9. Bartlett, N. W. *et al.* A 3D-printed, functionally graded soft robot powered by combustion. *Science* **349,** 161–165 (2015).
10. Peele, B. N., Wallin, T. J., Zhao, H. & Shepherd, R. F. 3D printing antagonistic systems of artificial muscle using projection stereolithography. *Bioinspir. Biomim.* **10,** 055003 (2015).
11. Mosadegh, B. *et al.* Integrated elastomeric components for autonomous regulation of sequential and oscillatory flow switching in microfluidic devices. *Nat. Phys.* **6,** 433–437 (2010).
12. Wehner, M. *et al.* Pneumatic energy sources for autonomous and wearable soft robotics. *Soft Robot.* **1,** 263–274 (2014).
13. Wu, W., Deconinck, A. & Lewis, J. A. Omnidirectional printing of 3D microvascular networks. *Adv. Mater.* **23,** H178–H183 (2011).
14. Muth, J. T. *et al.* Embedded 3D printing of strain sensors within highly stretchable elastomers. *Adv. Mater.* **26,** 6307–6312 (2014).
15. Palleau, E., Morales, D., Dickey, M. D. & Velev, O. D. Reversible patterning and actuation of hydrogels by electrically assisted ionoprinting. *Nat. Commun.* **4,** 2257 (2013).
16. Ionov, L. Biomimetic hydrogel-based actuating systems. *Adv. Funct. Mater.* **23,** 4555–4570 (2013).
17. Anderson, I. A., Gisby, T. A., McKay, T. G., O'Brien, B. M. & Calius, E. P. Multi-functional dielectric elastomer artificial muscles for soft and smart machines. *J. Appl. Phys.* **112,** 041101 (2012).
18. Brown, E. *et al.* Universal robotic gripper based on the jamming of granular material. *Proc. Natl Acad. Sci. USA* **107,** 18809–18814 (2010).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

19. Suzumori, K. Elastic materials producing compliant robots. *Robot. Auton. Syst.* **18,** 135–140 (1996).
20. Shepherd, R. F., Stokes, A. A., Nunes, R. M. D. & Whitesides, G. M. Soft machines that are resistant to puncture and that self seal. *Adv. Mater.* **25,** 6709–6713 (2013).
21. Suzumori, K., Koga, A., Kondo, F. & Haneda, R. Integrated flexible microactuator systems. *Robotica* **14,** 493–498 (1996).
22. Gorissen, B., De Volder, M., De Greef, A. & Reynaerts, D. Theoretical and experimental analysis of pneumatic balloon microactuators. *Sens. Actuat. A* **168,** 58–65 (2011).
23. Shepherd, R. F. *et al.* Using explosions to power a soft robot. *Angew. Chem. Int. Ed.* **52,** 2892–2896 (2013).
24. Whitesides, G. M. What comes next? *Lab Chip* **11,** 191–193 (2011).
25. Mosadegh, B., Tavana, H., Lesher-Perez, S. C. & Takayama, S. High-density fabrication of normally closed microfluidic valves by patterned deactivation of oxidized polydimethylsiloxane. *Lab Chip* **11,** 738–742 (2011).
26. Barnes, H. A. Thixotropy—a review. *J. Non-Newton. Fluid Mech.* **70,** 1–33 (1997).
27. Kolesky, D. B. *et al.* 3D bioprinting of vascularized, heterogeneous cell-laden tissue constructs. *Adv. Mater.* **26,** 3124–3130 (2014).
28. Randall, G. C. & Doyle, P. S. Permeation-driven flow in poly(dimethylsiloxane) microfluidic devices. *Proc. Natl Acad. Sci. USA* **102,** 10813–10818 (2005).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank D. Gessel, G. Leyh, M. Pauline, N. W. Bartlett, M. A. Skylar-Scott, T. J. Ober and J. T. Muth for their comments and discussions. We thank L. K. Sanders for assistance with photography and videography and J. C. Weaver for assistance with electron microscopy. We acknowledge support from the National Science Foundation through the Harvard MRSEC (grant no. DMR-1420570) and the Wyss Institute for Biologically Inspired Engineering. This work was performed in part at the Harvard University Center for Nanoscale Systems (CNS), a member of the National Nanotechnology Coordinated Infrastructure Network (NNCI), which is supported by the National Science Foundation under NSF award no. 1541959. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. R.L.T. acknowledges support from a National Science Foundation Graduate Research Fellowship and J.A.L. acknowledges support from the National Security Science and Engineering Faculty Fellowship.

Author Contributions M.W., R.L.T., J.A.L. and R.J.W. conceived the experimental work; M.W. and R.L.T. led the experiments with assistance from D.J.F. and B.M.; M.W., R.L.T., J.A.L. and R.J.W. contributed to data analysis and interpretation and wrote the paper. All authors provided feedback.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.J.W. (rjwood@seas.harvard.edu) and J.A.L. (jalewis@seas.harvard.edu).

Reviewer Information *Nature* thanks B. Mazzolai and the other anonymous reviewer(s) for their contribution to the peer review of this work.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

Soft controller fabrication. Soft controllers are fabricated from Sylgard 184 PDMS (Dow Corning Corp.) using soft lithography moulding and bonding techniques. First, a mould was fabricated on a silicon wafer using SU-8 negative photoresist (Microchem Corp.). SU-8 3050 photoresist was used to achieve 100-μm film thickness. Baking, exposing and developing steps were performed in accordance with product specifications in the product datasheet. The completed wafer is placed in a Petri dish to form a competed mould assembly.

Soft controllers consist of an upper mould, a lower mould and an intermediate thin film. The upper and lower moulds are made on one wafer to ease fabrication. PDMS is poured into the mould assembly to a height of 1 mm. Separately, PDMS is spin-coated onto a wafer at 1,500 r.p.m. for 60 s for a film thickness of 35 μm. After curing at 90 °C for 20 min, PDMS forms are removed from the moulds and holes are punched at all inlets and outlets. The upper layer is bonded to the wafer-adhered thin film after exposing to oxygen plasma at 35 W for 20 s in a Deiner Pico plasma system (Deiner Electronic GmbH). Holes are punched in the thin film, masks are placed as described in ref. 25, and the lower layer is bonded to the thin film using the plasma recipe above.

Ink and matrix materials. Two inks—a 'fugitive ink' and a 'catalytic ink'—are formulated for EMB3D printing. The fugitive ink is prepared by adding 27 wt% gel of Pluronic F127 to ice-cold, deionized, ultra-filtrated (DIUF) water, followed by mixing in a planetary mixer for 5 min at 2,000 r.p.m., and storing at 4 °C. The fugitive ink is not used until the Pluronic F127 completely dissolves in solution. The ink is prepared for printing by loading the solution at 4 °C in a 3-cm^3 syringe barrel (EFD Nordson) and centrifuged at 3,000 r.p.m. for 5 min to degas. For EMB3D printing, the barrel of the fugitive ink is fitted with a stainless steel nozzle (0.15-mm inner diameter; EFD Nordson).

The catalytic ink is prepared by first synthesizing and then dissolving a diacrylated Pluronic (F127-DA) at 30 wt% concentration with a solution of Irgacure 2959 (at 0.5 wt%, BASF) in DIUF water at 4 °C. The F127-DA is synthesized under an inert nitrogen atmosphere by first adding 400 ml of dry toluene (Sigma-Aldrich Co.) to a three-neck flask fixed to a condenser with circulating cold water and magnetically stirred at 300 r.p.m. 70 g of Pluronic F127 (Sigma-Aldrich Co.) is then dissolved in the toluene after heating the solvent to 60 °C. After the solution is allowed to cool to room temperature, triethylamine (5.6 g, Sigma-Aldrich Co.) is added to the solution, followed by the drop-wise addition of acryloyl chloride (5 g, Sigma-Aldrich Co.) with continued stirring, both at a molar ratio of 10:1 with the Pluronic F127. The reaction mixture is stirred overnight and maintained in the inert atmosphere. The diacrylated Pluronic F127 (F127-DA) product is then filtered from the yellow triethylammonium hydrochloride by-product and precipitated from the filtered solution with hexane (Sigma-Aldrich Co.) at a 1:1 volume ratio. The F127-DA is obtained through a second filtration step and allowed to dry in a chemical hood for at least 24 h. This protocol is adapted from ref. 13. For each gram of this base F127-DA mixture, 100 mg of PEG-DA is added, and this solution is mixed in a planetary mixer for 1 min at 2,000 r.p.m. and degassed for 3 min at 2,200 r.p.m. This mixture is then stored in the dark at 4 °C. Finally, 5 w/w% Pt black (Sigma-Aldrich Co.) is added to this base solution at 4 °C and mixed in a planetary mixer for 5 min at 2,000 r.p.m. The Pt-filled F127-DA physically gels during mixing, facilitating loading into a UV-blocking 3-cm^3 syringe barrel (EFD Nordson) for printing. This catalytic ink is freshly prepared for each print session, as the Pt black slowly cross-linked the acrylate moieties present in the ink. After EMB3D printing, the catalytic ink is cross-linked for 15 min at 18 mW cm^{-2} under a UV source (Omnicure EXFO). For EMB3D printing, the syringe barrel housing this ink is fitted with a stainless steel nozzle (0.33-mm inner diameter; EFD Nordson).

Two matrix materials are developed for fabricating fully soft robots. The first matrix, referred to as the 'body matrix', is prepared by blending two silicone-based materials: Sylgard 184 and SE 1700 (Dow Corning Corp.). SE 1700 is a silicone elastomer paste that contains fumed silica nanoparticles. Sylgard 184 PDMS is used to dilute SE 1700 to achieve the desired rheological response for embedded 3D printing. After exploring several blends, we found that the optimal body matrix is composed of a 1:1 mass ratio of SE 1700 (4:1 ratio of base to hardener) and Sylgard 184 (10:1 ratio of base to hardener). This matrix is prepared by mixing the blend in a planetary mixer at 2,000 r.p.m. for 3 min with degassing at 2,200 r.p.m. for 2 min. The second matrix, referred to as the 'fuel reservoir matrix', is prepared by mixing Part A Ecoflex 00-30 to Part B Ecoflex 00-30 (with 1.2 w/w% Slo-Jo Platinum Silicone Cure Retarder and 1.2 w/w% Thivex, Smooth-On Inc.) in a 1:1 ratio. The matrix is prepared in a planetary mixer at 2,000 r.p.m. for 1.5 min with degassing at 2,200 r.p.m. for 1 min.

Last, the 'fugitive plug' material used to prevent ingress of the body matrix material into the soft controller is prepared before printing by first synthesizing and then mixing a diacrylated Pluronic material (F127-DA) (at 30 wt% in a 0.5 wt% solution of Irgacure 2959 in deionized water) with F127 (at 30 wt% in deionized water) at a mass ratio of 1:4. The fugitive plug is stored in the dark at 4 °C in a syringe. When used, the fugitive plug material is allowed to physically gel before it is cross-linked for 3 min at 6 mW cm^{-2} under a UV source.

Rheological characterization. All rheological measurements are carried out using a controlled-stress rheometer (DHR-3, TA Instruments) equipped with a 40-mm-diameter, 2° cone and plate geometry. In all experiments, the fugitive and catalytic inks are equilibrated at room temperature for 1 min before testing; the fuel reservoir and body matrix materials are equilibrated for 20 min and 10 min, respectively, to simulate the times at which octobot printing began with each material. Shear storage moduli are measured as a function of shear stress at a frequency of 1 Hz.

The body matrix materials are characterized by both flow sweep and flow ramp tests to determine their rheological response (Extended Data Fig. 2). In addition, three-phase modulus recovery tests are carried out to quantify the recovery time of the body matrix stiffness after applying a shear stress that exceeds the equilibrium yield stress, $\tau_{y,0}$ (Extended Data Fig. 3). In the first set of experiments, flow sweeps from low (10^{-2} s^{-1}) to high (10^{2} s^{-1}) shear rates are carried out immediately followed by flow ramps from high to low shear rates. In the second set of experiments, shear storage (G') and loss (G'') moduli are measured during three phases of applied shear stresses (at 1-Hz frequency): 1 Pa for 3 min; 100 Pa for either 1 s, 10 s or 100 s; and 1 Pa for 30 min. We defined their thixotropic recovery time as the instant $G' = G''$, or when $\tan(\delta) = G''/G' = 1$, where δ is the phase angle.

Actuator characterization. Actuators are printed into special actuator characterization moulds by EMB3D printing and then auto-evacuate. To prepare them for characterization, they are first released from mould assembly and then a 1-mm hole is created with a biopsy punch (Miltex Inc.), which serves as the air inlet. Finally, the actuator is pressurized slightly to ensure inflation. Each actuator design is tested for angular displacement (that is, the actuator is allowed to deflect unconstrained and the total displacement angle is measured) and blocked force (that is, the actuator is constrained from deflection and resultant force is measured; see Extended Data Fig. 7). For each actuator, break-in testing consists of five cycles, in which actuator air pressure is slowly (over about 30 s) ramped up to the pressure set point, then slowly (over about 30 s) ramped down to ambient. Pressure set point for the first cycle is P_0 and the set point for all subsequent cycles is P_1 (Extended Data Table 1). Data acquisition consists of five additional cycles for each actuator, in which air pressure is cycled as above to pressure set point P_1.

To characterize their angular displacement, actuators are plumbed with regulated compressed air and mounted vertically between a matte black background and a Sony NEX3 digital camera for video data acquisition. Actuators are pressurized with five break-in cycles as described above, followed by five data-acquisition cycles. As above, the first break-in cycle is to P_0 and all subsequent break-in and data-acquisition cycles are to P_1. Video data are analysed using the ImageJ image analysis platform (NIH.gov) to obtain the bend angle versus pressure for each actuator.

For blocked-force characterization, individual actuators are mounted on a fixed platform beside an Instron model 5544 materials testing frame (Illinois Tool Works Inc.). The actuator is lowered until just above the force sensor portion of the testing frame and the actuator is plumbed with regulated compressed air. Actuators typically behave differently upon the initial few actuations versus subsequent actuations, owing to the Mullins effect[29]. Each actuator therefore receives five break-in cycles before data acquisition. Air pressure and actuator force data are recorded on the Instron testing frame data acquisition system at 100-ms intervals.

Mould fabrication. Octobot moulds are fabricated inside a CNC machined acetal mould equipped with two locating pins to mount the soft controller. Their desired shape is modelled in SolidWorks (Dessault Systemes SOLIDWORKS Corp.). A negative mould is modelled and output in Parasolid format for file transfer. MasterCAM (CNC Software, Inc.) is used to develop all machining tool paths and to export the final G-code for final fabrication. Blanks (12.7 cm × 7.6 cm × 2.54 cm) were cut from black acetal (Delrin) (McMaster Carr). Acetal is used, owing to its dimensional stability, and 2.54-cm-thick stock is chosen to prevent warping during machining and repeated octobot curing cycles. Octobot moulds are produced by CNC milling on a HAAS OM-2A vertical machining centre (HAAS Automation Inc.). 1-mm dowel pins are pressed into drilled holes for controller mounting.

Soft robot assembly. A custom-designed, multi-material 3D printer (ABL 10000, Aerotech Inc.) with four independently z-axis addressable ink reservoirs is used to pattern fugitive and catalytic inks within the octobot matrices[27]. All G-Code for printing is generated from Python-based software (MeCode, developed by J. Minardi). Prior to EMB3D printing, Ecoflex 30 (Smooth-On, Inc.) is first prepared with 1 wt% Slo-Jo and 0.25 wt% Thivex (both with respect to Part A) by mixing in a Thinky planetary mixer for 1.5 min with a 1-min degas cycle. This uncured Ecoflex 30 is cast into the actuator layers of the octobot mould and degassed in a vacuum chamber for 3 min. A glass slide is used to remove excess material and create smooth surfaces that will ultimately become the extensible layers of the actuators. The moulds are then placed in a 90 °C oven for 30 min to cure the Ecoflex, removed, and trimmed of excess material as necessary.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

A soft controller is then loaded onto the press-fit pins placed in the printing mould with the polyimide (Kapton) tape still adhered. Registration coordinates and print heights are then taken from the cured Ecoflex layers in the actuators and in all inlets of the soft controller; these are essential for EMB3D printing and provided to the custom print software. The fuel reservoir and body matrix materials are prepared as described previously. While the body matrix material is mixing, the fuel reservoir matrix is deposited in the fuel reservoir region of the printing mould. It is then degassed for 3 min to ensure no trapped gas is present. Excess bubbles in the uncured fuel reservoir matrix are removed with a pipettor. The non-gelled, chilled fugitive plug is then filled throughout the soft controller via injection through the inlets. While the fugitive plug is still in the liquid state, it is briefly degassed in a vacuum chamber. The fugitive plug material is then allowed to physically cross-link, excess gel is scraped from the top of the tape, the tape is removed, and the fugitive plug is photo-cross-linked with a UV source at 6 mW cm^{-2} for 3 min. After the gels are cross-linked, the body matrix is cast within the mould, covering the fuel reservoir matrix and the fugitive plug-filled soft controller and degassed for 1–3 min *in situ*. Again, excess bubbles are removed with a pipettor, excess material is scraped off and away from the mould with a glass slide, and EMB3D printing of the fugitive and catalytic inks begins. After printing, the entire mould is cured at 18 mW cm^{-2} for 15 min to crosslink the catalytic ink. The mould is then transferred to a 90 °C oven, where the matrix materials cross-link. The octobot is removed from the mould and kept at 90 °C for 4 days to facilitate auto-evacuation of the inks.

After auto-evacuation, the octobot is release-cut from the surrounding matrix material using a CO_2 laser (Universal Laser Systems) and cleaned with isopropyl alcohol and water. Sylgard 184 PDMS (Dow Corning Corp.) is poured into the open cavity of the octobot above the soft controller to a height of 1.5 mm and cured at 90 °C for 20 min. A 1-mm biopsy punch (Miltex Inc.) is used to punch holes through the newly poured PDMS layer and into the fuel inlets. Dyed water is injected into these holes to inflate the fuel tanks, flow through the system and insure proper bot function. Holes are punched in the downstream vent orifices to allow the water to vent from the system.

The octobot is loaded into an acrylic tank outfitted with a backlight to highlight coloured fuel as it flows through the system. Aqueous hydrogen peroxide (90 wt%, HTP grade, Peroxychem) is diluted to 50 wt% and samples dyed red and blue are filled into two syringes prepared with this liquid fuel mixture. The syringes are loaded onto a syringe pump, and connected to the octobot via 1-mm-diameter silicone rubber tubing. Water is flowed into the acrylic tank to wash away dye in the octobot exhaust stream and drained into a nearby sink. The syringe pump flows fuel at a rate of 3 ml min^{-1} (each syringe) into the octobot for 10 s. The silicone rubber tubing is removed with tweezers from the octobot, which is allowed to operate untethered. The octobot alternates actuation until fuel pressure is insufficient to switch the oscillator and alternating actuation ceases.

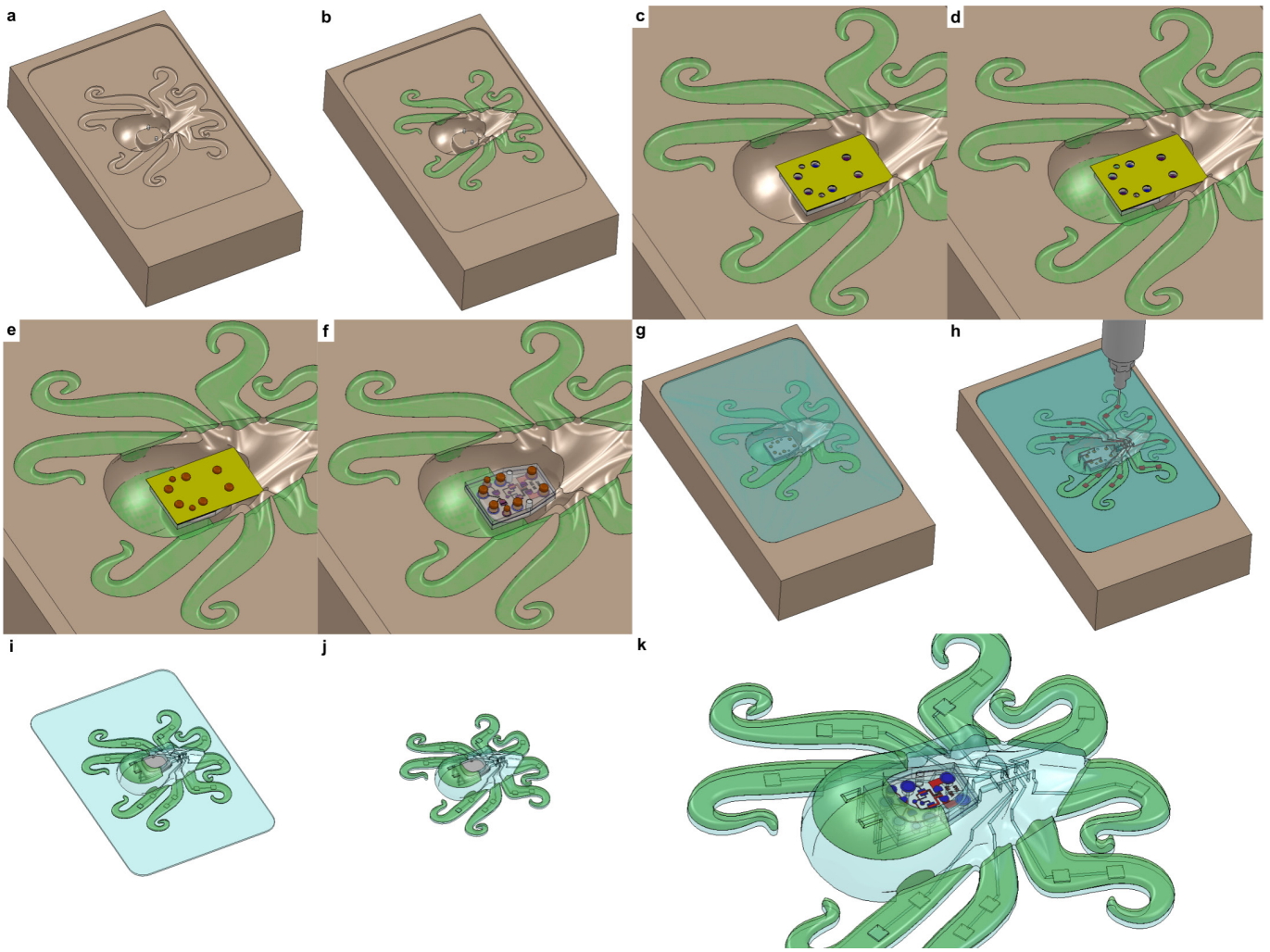
Imaging and videography. Photographs and supporting videos are acquired with a digital SLR camera (Canon EOS 5D Mark II, Canon USA Inc.) and a 4K video (Blackmagic Production 4K, Blackmagic Design). Photos are cropped using Inkscape vector graphics editor (http://www.inkscape.org) and video sequences are clipped from raw footage and exported using iMovie (Apple Corp.; titles are added using Premiere Pro, Adobe Systems, Inc.). All print parameter measurements and images of EMB3D printed features in octobots are taken with a digital zoom microscope (VHX-2000, Keyence). Their mean values and standard deviations are determined from three samples printed at each print speed of interest.

Considerations for future microfluidic controllers, logic and actuation. The octobot represents a minimal soft robotic system that demonstrates our integrated design and fabrication strategy. In our self-contained microfluidic controllers, system scaling is limited by fuel flow rate, on-board fuel supply, downstream decomposition/expansion and actuation-network design. At 50% concentration by weight, 1 g of aqueous hydrogen peroxide expands to approximately 200 ml of gas under ambient pressure and temperature conditions[12]. Our oscillator is designed to operate at 40 μl min^{-1} per channel (80 μl min^{-1} total), and our actuators inflate with approximately 0.2 ml at 50 kPa, the equivalent of 0.3 ml at ambient pressure. Hence, with each channel inflating four actuators, our current design has a theoretical maximum oscillation rate of 5.5 switches per minute. Although the controller and actuators could be redesigned for increased performance, any system scaling would require careful balancing of fuel supply, flow rate and actuator requirements. Alternative actuator designs are possible (see Extended Data Fig. 7), but our current actuators produce 0.04 N; therefore, two actuators would theoretically be sufficient to lift the 7-g robot.

On the basis of these demonstrated capabilities, we anticipate that more sophisticated microfluidic control systems and logic devices could be readily incorporated within these printed and moulded robots. For example, fluidic versions of electric logic gates have been reported, including NAND/NOR, AND/OR and XOR/XNOR[30–32], and flip-flops and gain valves[33–35]. These complex systems are based on well-established, electrical design rules. Implementation of these systems would enable more complex actuation strategies, such as multi-degree-of-freedom actuators in which planned limb motion would prescribe a true gait with aerial and ground phases to lift and propel the soft robot. Alternatively, actuators could be designed to take advantage of material elasticity, in which actuation performs flexion and abduction, and passive limb elasticity provides extension and adduction. One could even envision an actuation strategy in which pneumatic channels act as sensors, providing true closed-loop feedback to the controller.

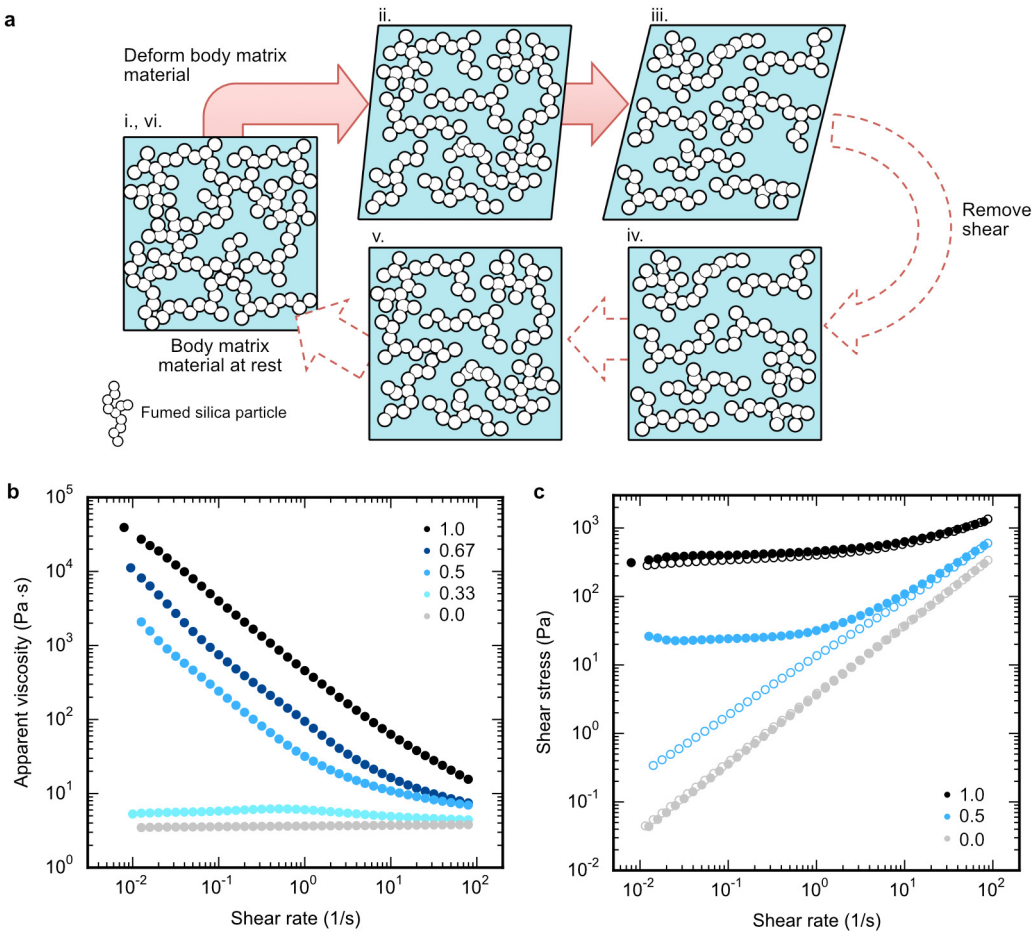
29. Mullins, L. Softening of rubber by deformation. *Rubber Chem. Technol.* **42,** 339–362 (1969).
30. Jensen, E. C., Grover, W. H. & Mathies, R. A. Micropneumatic digital logic structures for integrated microdevice computation and control. *J. Microelectromech. Syst.* **16,** 1378–1385 (2007).
31. Rhee, M. & Burns, M. A. Microfluidic pneumatic logic circuits and digital pneumatic microprocessors for integrated microfluidic systems. *Lab Chip* **9,** 3131–3143 (2009).
32. Toepke, M. W., Abhyankar, V. V. & Beebe, D. J. Microfluidic logic gates and timers. *Lab Chip* **7,** 1449–1453 (2007).
33. Groisman, A., Enzelberger, M. & Quake, S. R. Microfluidic memory and control devices. *Science* **300,** 955–958 (2003).
34. Weaver, J. A., Melin, J., Stark, D., Quake, S. R. & Horowitz, M. A. Static control logic for microfluidic devices using pressure-gain valves. *Nat. Phys.* **6,** 218–223 (2010).
35. Prakash, M. & Gershenfeld, N. Microfluidic bubble logic. *Science* **315,** 832–835 (2007).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



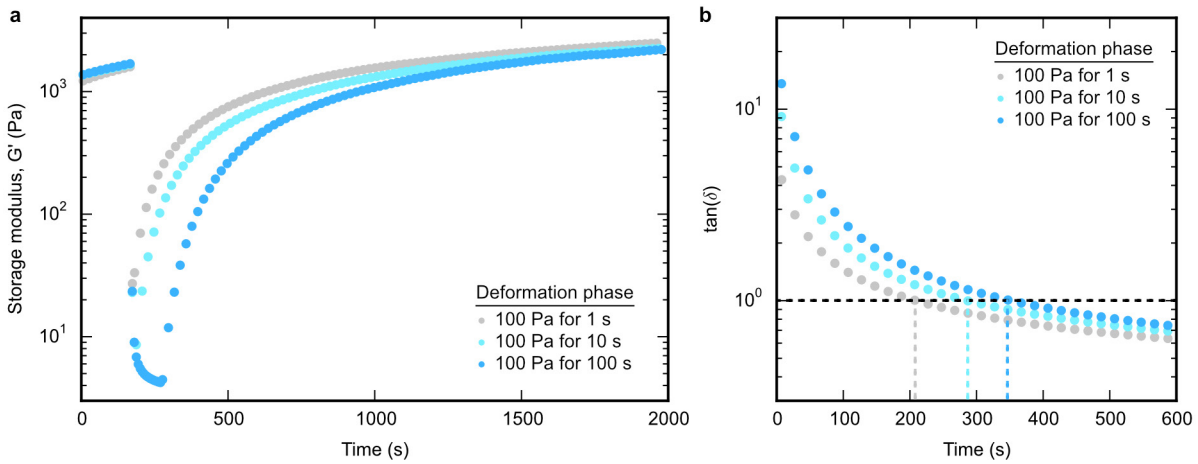
Extended Data Figure 1 | EMB3D printing of an octobot. a, An EMB3D printing mould is machined from acetal. **b**, The hyperelastic layers needed for actuation are cast and cross-linked in the actuator regions of the mould. **c**, A soft controller protected with a polyimide tape mask is loaded onto the pins of the EMB3D printing mould. **d**, The fuel reservoir matrix material is carefully loaded into the fuel reservoir area of the mould and degassed under vacuum. **e**, Liquefied fugitive plug material is manually loaded into the soft controller via the inlets and briefly degassed. **f**, The protective tape is removed after the fugitive plug material physically gels, and the fugitive plug is photo-cross-linked. **g**, The body matrix material is cast into the mould and degassed. **h**, Any excess body matrix material is removed with a squeegee step, EMB3D printing begins, and the entire mould and EMB3D-printed materials are placed in a 90 °C oven to cross-link. **i**, After 2 h, the cross-linked octobot is removed from its mould and kept at 90 °C for a total of 4 days to ensure complete auto-evacuation of the aqueous fugitive inks. **j**, Before operation, excess body matrix material is removed via laser cutting. **k**, The final octobot, shown here in a close-up view, is prepared for operation.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



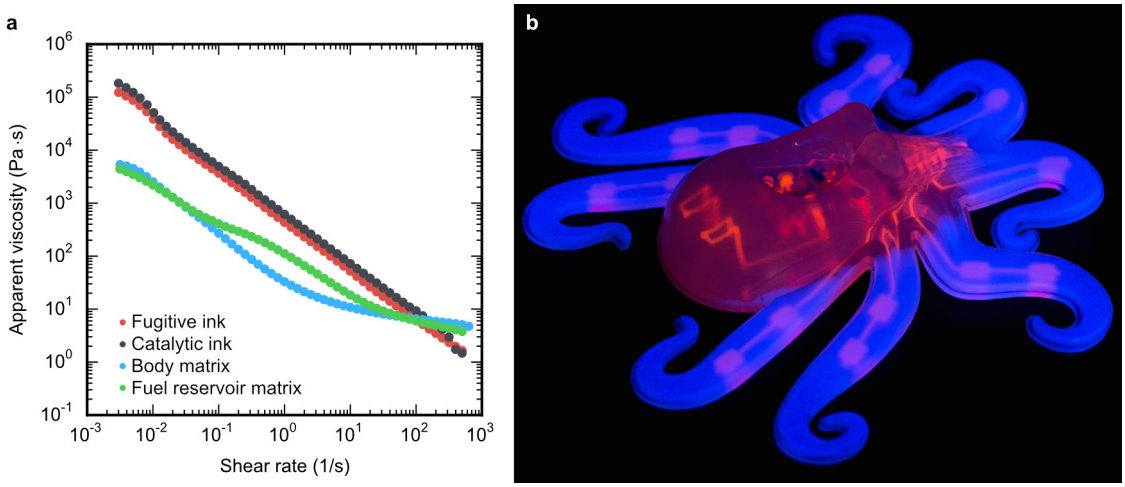
Extended Data Figure 2 | Rheological properties of the body matrix. **a**, Schematic illustration of the behaviour of the body matrix during the EMB3D printing process. (i) When the body matrix is at rest, the fumed silica fillers within the silicone material form a percolated network, giving rise to its equilibrium, at-rest shear yield stress, $\tau_{y,0}$. (ii) As the nozzle travels through the matrix during EMB3D printing, the matrix is yielded and the percolated filler network is disrupted, decreasing the apparent yield stress of the matrix material, $\tau_{y,t}$. (iii) Sufficient deformation can completely disrupt the fumed silica microstructure and completely eliminate the yield stress of the matrix material ($\tau_{y,t} \rightarrow 0$ Pa). (iv) The fumed silica network does not immediately recover when it returns to a quiescent state. (v) Over time, the network slowly restructures to (vi) its equilibrium microstructure, and $\tau_{y,t} \rightarrow \tau_{y,0}$. **b**, **c**, Log–log plots of apparent viscosity (**b**) and corresponding shear stress (**c**) versus shear rate for various PDMS matrix formulations, which are prepared by blending Sylgard 184 (10:1 ratio of base to hardener) and SE 1700 (4:1 ratio of base to hardener) at various mass fractions. The formulations are listed by the weight ratio of SE 1700 used (0.0,0.33, 0.5, 0.67 and 1.0). Closed and open circles in **c** represent measurements taken during the flow sweep and flow ramps of the thixotropic loop studies, respectively. The final body matrix, formulated from the 50 wt% SE 1700 blend, shows clear thixotropic behaviour and a substantial decrease in yield stress upon yielding. Blends with higher concentrations of filler particles show diminished thixotropic behaviour, and the yield stress is not eliminated during nozzle translation. Consequently, crevices or air pockets form during printing, with matrix materials possessing higher concentrations of fumed silica.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



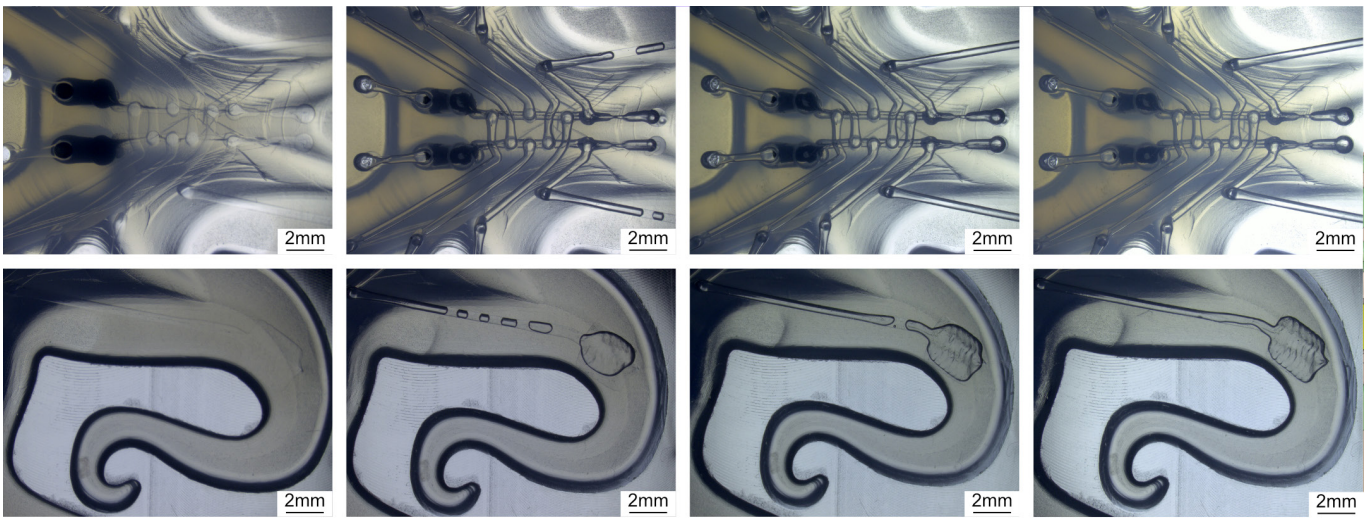
Extended Data Figure 3 | Storage modulus recovery of the body matrix after yielding. a, A plot of storage modulus (G') as a function of time illustrates how the modulus of the body matrix recovers during three-phase thixotropy tests. After a probe phase, a shear stress of 100 Pa is applied for varying times during a deformation phase, resulting in temporary fluidization of the matrix material. During the recovery phase, the modulus increases over time. **b**, The ratio of the loss modulus to the storage modulus, $G''/G' = \tan(\delta)$, is plotted as a function time for each of the recovery phases measured in **a**. The onset of recovery of the yield stress of the body matrix material—and the onset of fumed silica filler percolation in a recovering matrix material—is assumed to be the moment $G' = G''$ or $\tan(\delta) = 1$ (horizontal dashed line). Therefore, the 'recovery time' of the body matrix material (indicated by the vertical dashed lines) is approximately the time at which $\tan(\delta) = 1$ after deformation. Because the momentary deformation incurred by nozzle translation through a discrete volume of matrix material during EMB3D printing happens within a time period shorter than 1 s and with a magnitude of less than 100 Pa, the thixotropic recovery time of the body matrix material is less than 200 s—the approximate time it takes the body matrix to recover after being sheared by a 100-Pa stress for 1 s.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



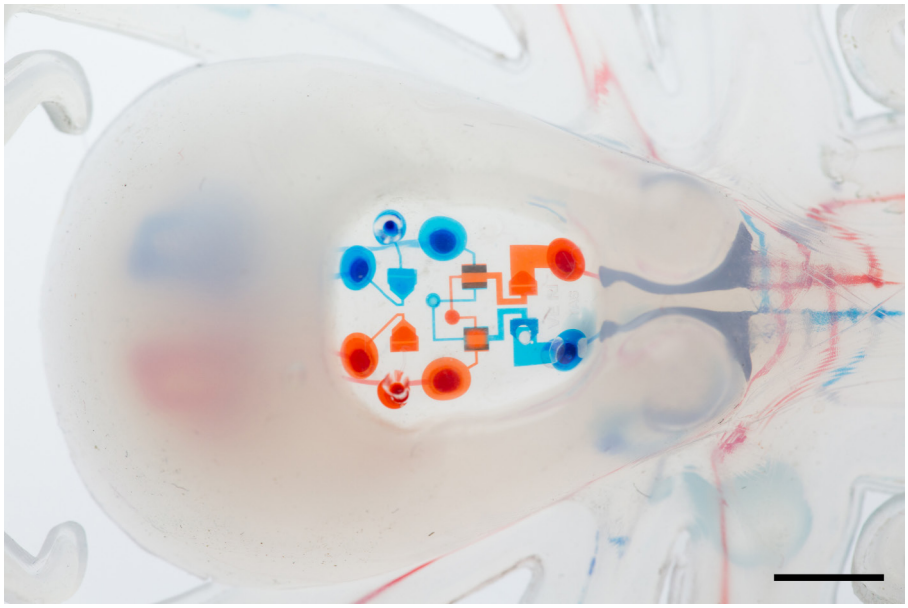
Extended Data Figure 4 | Rheological and printing behaviour of inks and matrix materials. a, A log–log plot of apparent viscosity as a function of shear rate for the fugitive ink (red), catalytic ink (black), body matrix material (blue) and fuel reservoir matrix material (green). **b**, An octobot with fluorescently dyed fugitive inks (red, not auto-evacuated) and hyperelastic actuator layers (blue) fabricated by moulding and EMB3D printing.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



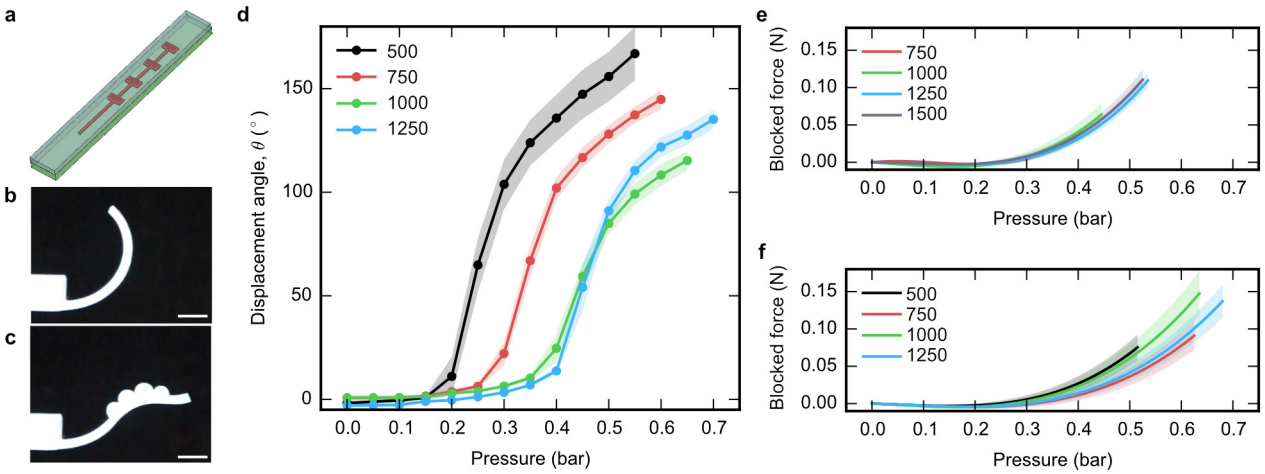
Extended Data Figure 5 | Auto-evacuation of the fugitive and catalytic inks. Photographs of the reaction chambers of an octobot showing the upstream portions of the actuator networks (top) and a one-pad actuator (bottom) at various times, *t*. These photographs reveal the auto-evacuation of the fugitive and catalytic inks, which leaves behind open channels that serve as mesofluidic features.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



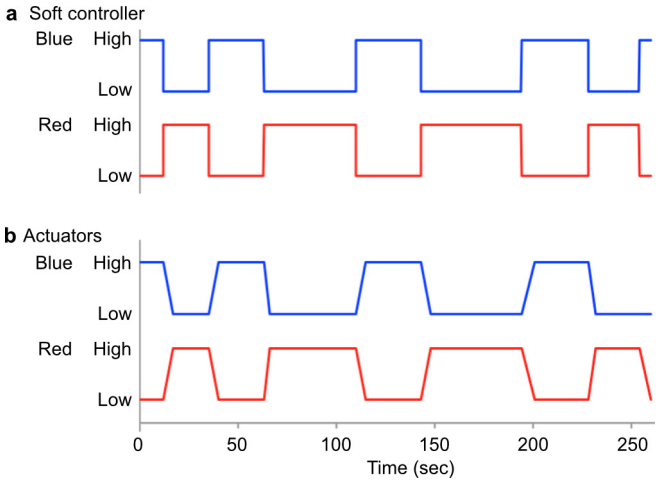
Extended Data Figure 6 | Infilling the soft controller from the fuel inlets. Water (with red or blue dye) is introduced into the fuel reservoir via the fuel inlets. Continuity between the fuel reservoirs, soft controller and downstream EMB3D-printed components is possible because of the fugitive plugs, which auto-evacuate along with the EMB3D-printed inks. Scale bar, 5 mm.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 7 | Characterization of EMB3D-printed actuators. a, CAD model of a four-bladder actuator design. Other than bladder number, the design is similar to the actuators illustrated in Fig. 4a. **b**, EMB3D-printed actuator is shown before inflation. **c**, Actuator inflated to working pressure. Scale bars in **b** and **c**, 5 mm. **d**, Pressure versus displacement curves for four-bladder actuators with varying hyperelastic-layer thickness, *h* (see legend; values in micrometres). The data points reported are the mean inflation displacement angles for two representative actuators over five inflation cycles and the shading indicates 95% confidence intervals. **e**, **f**, Blocked force versus pressure curves for two-bladder (**e**) and four-bladder (**f**) actuators of varying hyperelastic-layer thickness, *h* (see legends; values in micrometres). The lines shown are third-degree polynomial fits of data collected from five inflation–deflation cycles of representative actuators. The shaded regions indicate 95% confidence intervals. Detailed procedures for characterizing actuator performance are provided in Methods.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 8 | Autonomous switching between actuating states during octobot operation. a, Switching in the soft controller was tracked with time according to the octobot operation recorded in Supplementary Video 5. **b**, The corresponding inflation of actuators associated with the blue and red and actuation states are also reported with time.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 1 | Break-in and working gauge pressures for printed actuators

	Two bladder actuators		Four bladder actuators	
Hyperelastic layer thickness (μm)	P_0 (Bar)	P_1 (Bar)	P_0 (Bar)	P_1 (Bar)
500	0.35	0.3	0.6	0.55
750	0.4	0.35	0.65	0.6
1000	0.45	0.4	0.7	0.65
1250	0.5	0.45	0.75	0.7
1500	0.55	0.5	0.8	0.75

Before characterizing an actuator, break-in testing consisted of five cycles, in which actuator air pressure is slowly (over about 30 s) ramped up to the pressure set point, then slowly (over about 30 s) ramped down to ambient pressure. The pressure set point for the first cycle is P_0 and the set point for all subsequent cycles is P_1. Data acquisition consists of five additional cycles for each actuator, in which air pressure is cycled as above to pressure set point P_1.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

LETTER

doi:10.1038/nature19092

Biodiversity at multiple trophic levels is needed for ecosystem multifunctionality

Santiago Soliveres[1], Fons van der Plas[1,2], Peter Manning[1,2], Daniel Prati[1], Martin M. Gossner[3,4], Swen C. Renner[5,6], Fabian Alt[7], Hartmut Arndt[8], Vanessa Baumgartner[9], Julia Binkenstein[10], Klaus Birkhofer[11], Stefan Blaser[1], Nico Blüthgen[12], Steffen Boch[1,13], Stefan Böhm[5], Carmen Börschig[14], Francois Buscot[15,16], Tim Diekötter[17], Johannes Heinze[18,19], Norbert Hölzel[20], Kirsten Jung[21], Valentin H. Klaus[20], Till Kleinebecker[20], Sandra Klemmer[15], Jochen Krauss[22], Markus Lange[3,4,23], E. Kathryn Morris[24,25], Jörg Müller[18], Yvonne Oelmann[7], Jörg Overmann[9], Esther Pašalić[3,4], Matthias C. Rillig[19,25], H. Martin Schaefer[26], Michael Schloter[27], Barbara Schmitt[1], Ingo Schöning[3,23], Marion Schrumpf[23], Johannes Sikorski[9], Stephanie A. Socher[28], Emily F. Solly[23,29], Ilja Sonnemann[30], Elisabeth Sorkau[7], Juliane Steckel[22], Ingolf Steffan-Dewenter[22], Barbara Stempfhuber[27], Marco Tschapka[21,31], Manfred Türke[3,4,16,32], Paul C. Venter[8], Christiane N. Weiner[12], Wolfgang W. Weisser[3,4], Michael Werner[22], Catrin Westphal[14], Wolfgang Wilcke[33], Volkmar Wolters[34], Tesfaye Wubet[15,16], Susanne Wurst[30], Markus Fischer[1,2,13] & Eric Allan[1,35]

Many experiments have shown that loss of biodiversity reduces the capacity of ecosystems to provide the multiple services on which humans depend[1,2]. However, experiments necessarily simplify the complexity of natural ecosystems and will normally control for other important drivers of ecosystem functioning, such as the environment or land use. In addition, existing studies typically focus on the diversity of single trophic groups, neglecting the fact that biodiversity loss occurs across many taxa[3,4] and that the functional effects of any trophic group may depend on the abundance and diversity of others[5,6]. Here we report analysis of the relationships between the species richness and abundance of nine trophic groups, including 4,600 above- and below-ground taxa, and 14 ecosystem services and functions and with their simultaneous provision (or multifunctionality) in 150 grasslands. We show that high species richness in multiple trophic groups (multitrophic richness) had stronger positive effects on ecosystem services than richness in any individual trophic group; this includes plant species richness, the most widely used measure of biodiversity. On average, three trophic groups influenced each ecosystem service, with each trophic group influencing at least one service. Multitrophic richness was particularly beneficial for 'regulating' and 'cultural' services, and for multifunctionality, whereas a change in the total abundance of species or biomass in multiple trophic groups (the multitrophic abundance) positively affected supporting services. Multitrophic richness and abundance drove ecosystem functioning as strongly as abiotic conditions and land-use intensity, extending previous experimental results[7,8] to real-world ecosystems. Primary producers, herbivorous insects and microbial decomposers seem to be particularly important drivers of ecosystem functioning, as shown by the strong and frequent positive associations of their richness or abundance with multiple ecosystem services. Our results show that multitrophic richness and abundance support ecosystem functioning, and demonstrate that a focus on single groups has led to researchers to greatly underestimate the functional importance of biodiversity.

Global change is causing species loss across many trophic groups[3,4], with potential effects on the services that ecosystems provide to humans[1,2]. The functional consequences of a decline in biodiversity across multiple trophic groups are hard to predict from studies focusing on single taxa, as the functional effects of different groups may complement or oppose each other[5,6,9,10]. The effects of the diversity of plants and microbes are complementary, maximizing rates of nutrient cycling[11]; plant and herbivore diversity, on the other hand, have opposing effects on biomass stocks[10,12,13]. Consequently, we know very little about the relative effect of changes in the diversity of different trophic groups on the provision of individual[2,5,6,9,13,14] or multiple (multifunctionality)[11,15] ecosystem services.

In addition to decreasing species richness, global change is altering the total abundance (total number of individuals or amount of biomass within communities) of multiple trophic groups[4]. Changes in

[1]Institute of Plant Sciences, University of Bern, Altenbergrain 21, 3013 Bern, Switzerland. [2]Senckenberg Gesellschaft für Naturforschung, Biodiversity and Climate Research Centre BIK-F, Senckenberganlage 25, 60325 Frankfurt, Germany. [3]Institute of Ecology, Friedrich-Schiller-University Jena, Dornburger Straße 159, D-07743 Jena, Germany. [4]Technische Universität München, Terrestrial Ecology Research Group, Department of Ecology and Ecosystem Management, School of Life Sciences Weihenstephan, Hans-Carl-von-Carlowitz-Platz 2, 85354 Freising, Germany. [5]Institute of Zoology, University of Natural Resources and Life Science, Gregor-Mendel-Straße 33, 1180 Vienna, Austria. [6]Smithsonian Conservation Biology Institute, National Zoological Park, 1500 Remount Road, Front Royal, Virginia 22630, USA. [7]Geocology, University of Tuebingen, Ruemelinstr. 19-23, 72070 Tuebingen, Germany. [8]University of Cologne, Institute for Zoology, Zülpicher Str. 47b, 50674 Cologne, Germany. [9]Leibniz Institute DSMZ German Collection of Microorganisms and Cell Cultures, Inhoffenstr. 7B, 38124 Braunschweig, Germany. [10]Chair of Nature Conservation and Landscape Ecology, Faculty of Environment and Natural Resources, University of Freiburg, Tennenbacher Straße 4, 79106 Freiburg, Germany. [11]Department of Biology, Lund University, Sölvegatan 35, 22362 Lund, Germany. [12]Ecological Networks, Biology, Technische Universität Darmstadt, Schnittspahnstr. 3, 64287 Darmstadt, Germany. [13]Botanical Garden, University of Bern, Altenbergrain 21, 3013 Bern, Switzerland. [14]Agroecology, Department of Crop Sciences, Georg-August University of Göttingen, Grisebachstr. 6, D-37077, Göttingen, Germany. [15]UFZ-Helmholtz Centre for Environmental Research, Department of Soil Ecology, Theodor-Lieser-Straße 4, 06120 Halle (Saale), Germany. [16]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, D-04103 Leipzig, Germany. [17]Department of Landscape Ecology, Kiel University, Olshausenstr. 75, D-24118 Kiel, Germany. [18]Biodiversity Research/Systematic Botany, University of Potsdam, Maulbeerallee 1, D-14469 Potsdam, Germany. [19]Berlin-Brandenburg Institute of Advanced Biodiversity Research (BBIB), D-14195 Berlin, Germany. [20]Institute of Landscape Ecology, University of Münster, Heisenbergstr. 2, 48149 Münster, Germany. [21]Institute of Evolutionary Ecology and Conservation Genomics, University of Ulm, Albert-Einstein-Allee 11, 89069 Ulm, Germany. [22]Department of Animal Ecology and Tropical Biology, Biocentre, University of Würzburg, Am Hubland, D-97074 Würzburg, Germany. [23]Max-Planck Institute for Biogeochemistry, Hans-Knoell-Str. 10, 07745 Jena, Germany. [24]Xavier University, Department of Biology, 3800 Victory Parkway, Cincinnati, Ohio 45207, USA. [25]Plant Ecology, Institut für Biologie, Freie Universität Berlin, Altensteinstr. 6, D-14195 Berlin, Germany. [26]Department of Ecology and Evolutionary Biology, Faculty of Biology, University of Freiburg, Hauptstraße 1, 79104 Freiburg i. Br., Germany. [27]Research Unit for Environmental Genomics; Helmholtz Zentrum München, Ingolstädter Landstr. 1, 85758 Oberschleissheim, Germany. [28]Department of Ecology and Evolution, Universität Salzburg, Hellbrunnerstrasse 34, 5020 Salzburg, Austria. [29]Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Zürcherstrasse 111, 8903 Birmensdorf, Switzerland. [30]Functional Biodiversity, Institute of Biology, Freie Universität Berlin. Königin-Luise-Str. 1-3. D-14195 Berlin, Germany. [31]Smithsonian Tropical Research Institute, Balboa, Panama. [32]Institute for Biology, Leipzig University, Johannisallee 21, D-04103 Leipzig, Germany. [33]Institute of Geography and Geoecology, Karlsruhe Institute of Technology (KIT), Reinhard-Baumeister-Platz 1, 76131 Karlsruhe, Germany. [34]Department of Animal Ecology, Justus Liebig University Giessen, Heinrich-Buff-Ring 26-32, D-35392 Giessen, Germany. [35]Centre for Development and Environment, University of Bern, Hallerstrasse, 10, 3012 Bern, Switzerland.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

abundance could mitigate or exacerbate the functional consequences of species loss[16,17] by influencing the ability of each trophic group to capture resources. However, studies normally focus on the effects of community evenness or of dominant species[18–20], whereas the simultaneous effects of changes in richness and total abundance on the functioning of ecosystems have been largely unexplored[6,16]. The relative importance of richness and abundance may depend on the function or service of interest. Total abundance could be a main driver of biogeochemical process rates (for example, biomass production[18], nutrient capture and cycling). By contrast, ecosystem services related to biotic interactions, such as pollination or pest control, could be predominantly driven by species richness[16]. Ecosystem services also depend on abiotic factors and, although experiments show that the effects of biodiversity loss on ecosystem functioning are as large as those of abiotic drivers[7,8], it is unclear whether species richness and abundance are similarly important in real-world ecosystems[6,14,21,22].

We adopted a multitrophic approach to evaluate relationships between biodiversity and multifunctionality in 150 real-world grasslands. We measured the richness and abundance of species in nine trophic groups: primary producers, above- and below-ground herbivores and predators, detritivores, soil microbial decomposers, plant symbionts, and bacterivores. These trophic groups comprised 4,600 plant, animal and microbial taxa, and were measured alongside 14 ecosystem variables (proxies for both functions and services, hereafter referred to as services). These are related to the four main types of ecosystem services[23]: provisioning (fodder production and quality), supporting (potential nitrification, phosphorus retention, root biomass and decomposition rate, mycorrhizal colonization and soil aggregate stability), regulating (soil carbon levels, pollinator abundance, pest control, resistance to pathogens) and cultural services (recreation benefits of flower cover and bird diversity). We fitted linear models to our data to test for both positive and negative relationships between the richness and abundance of species within the nine trophic groups and each ecosystem service, the four types of services (provisioning, supporting, regulating and cultural), and ecosystem multifunctionality[22] (see Methods). We accounted for potential confounding factors by performing our analyses on residuals, after controlling for variability in land-use intensity, soils and climate. We compared our results with models that included only plant-species richness, the most commonly used measure of biodiversity[21,24,25], and with models that included the richness and abundance of each individual trophic group. Additional analyses compared the amount of variance explained by, and the effect size (standardized slope) of, multitrophic richness and abundance with those of land-use intensity and environmental variables.

Effects on individual ecosystem services, service types, and multifunctionality were better predicted by changes in multitrophic richness and abundance than by those in the richness or abundance of any individual trophic group (Fig. 1 and Extended Data Fig. 1). The most parsimonious models included the richness and/or abundance of 3.14 ± 0.36 trophic groups (average $\pm$ s.e.m. across all 14 services) to predict the variation in each ecosystem service. These results remained when using raw data instead of environment-corrected residuals (Extended Data Fig. 2), different combinations of ecosystem services (Extended Data Fig. 3), and even when we accounted for well-established links between a predictor and service in our models (for example, plant cover versus biomass). Multitrophic richness had stronger and more positive relationships with the provisioning, regulating and cultural services than plant richness alone (Fig. 1 and Extended Data Fig. 1). For example, both plant and predator richness were related to high levels of pest control, suggesting that combined top-down and bottom-up effects of diversity[26] maximize the provision of this regulating service. Multitrophic richness also had a more positive effect than even the strongest positive-richness effect found across all individual trophic groups on the regulating and cultural services. The findings of our observational study were supported by a quantitative review of the few studies that manipulated the richness of more than one group. Our review showed that including the richness of a second trophic group increased the variance in ecosystem functioning by 14–96% for litter decomposition[14], biomass production[2,12,26], or the number of carbon sources used[5] (Extended Data Table 1). Collectively, our results show that high species richness in multiple trophic groups is necessary to maintain high levels of ecosystem functioning, particularly for regulating and cultural services.

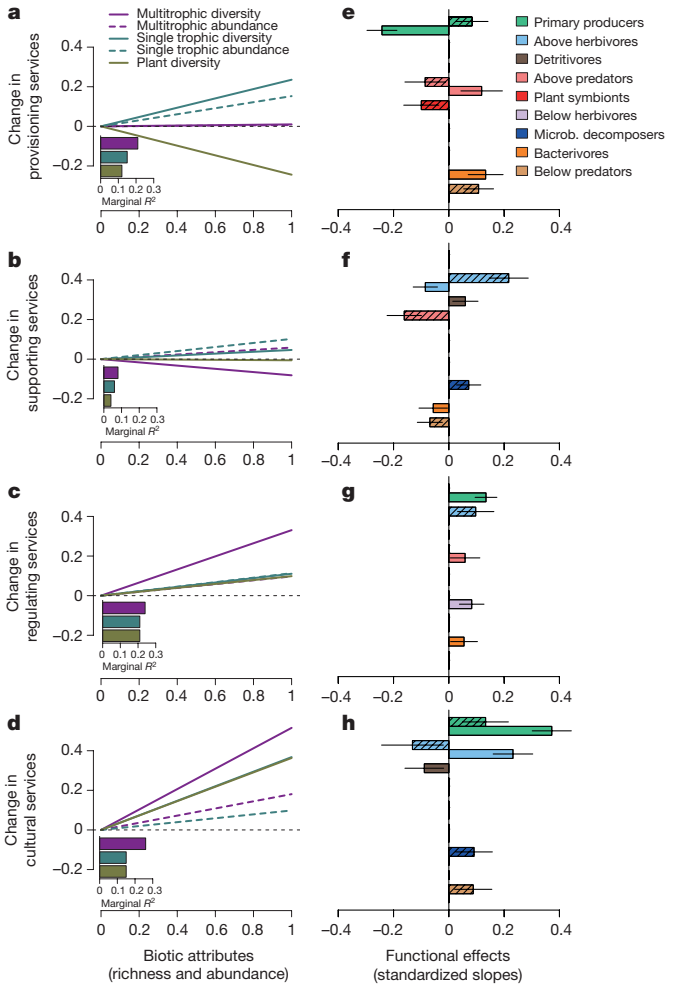


Figure 1 | Effects of multitrophic richness and abundance on grassland functioning. a–d, Variance explained after accounting for the influence of 'site' as random factor (marginal R^2, the equivalent of R^2 for mixed models) and standardized effects for each ecosystem service when models included abundance and richness of multiple or individual trophic groups. **e–h**, Standardized effects (mean $\pm$ s.e.m.) of richness and abundance (full and hatched bars, respectively) of individual trophic groups on each ecosystem service type. Ecosystem services types are plant biomass and forage quality (provisioning); potential nitrification, phosphorus retention, mycorrhizal colonization, soil aggregate stability, root biomass and decomposition (supporting); soil carbon, pollinator abundance, pest control and resistance to pathogens (regulating); flower cover and bird diversity (cultural).

Alongside multitrophic richness, the combined effect of a high multitrophic abundance strongly affected ecosystem functioning (according to the amount of variance explained and its effect size). Multitrophic abundance had positive effects on the provisioning and supporting services, but these were generally weaker than those found for the individual trophic group that had the strongest positive effect. This suggests that an abundance of some trophic groups can dampen the effect on ecosystem functioning induced by others. Figure 1, for instance, shows that a higher abundance of predators partially counteracted the positive effects of abundant herbivores on supporting services. Conversely, a high level of richness in a given trophic

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

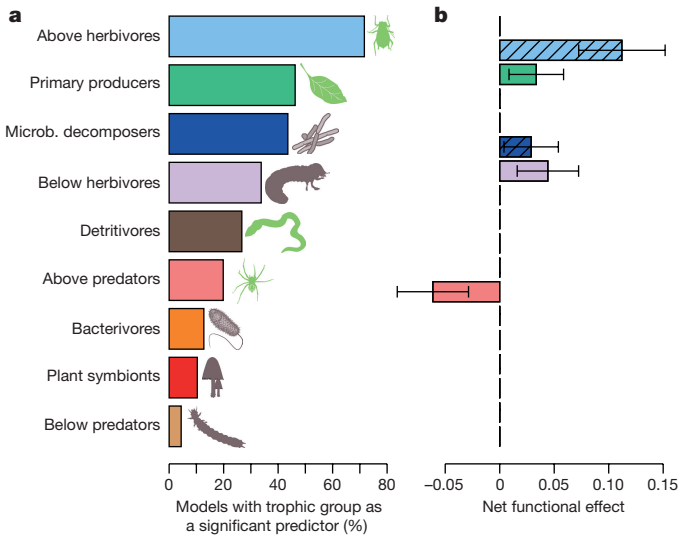


Figure 2 | Functional importance of multiple trophic groups. **a**, Proportion of the multifunctionality metrics (calculated using every possible combination of 1–9 services; $N = 501$; see Methods) in which the biotic attributes (richness and/or abundance) of each trophic group was included in the most parsimonious model. **b**, Functional effects (standardized slopes (mean ± s.e.m.) in the model fitted to all 14 services) of the richness (open bars) and abundance (hatched bars) of each group. Bars are shown only for the predictors included in the most parsimonious models. Green and brown cartoons indicate above- and below-ground trophic groups, respectively.

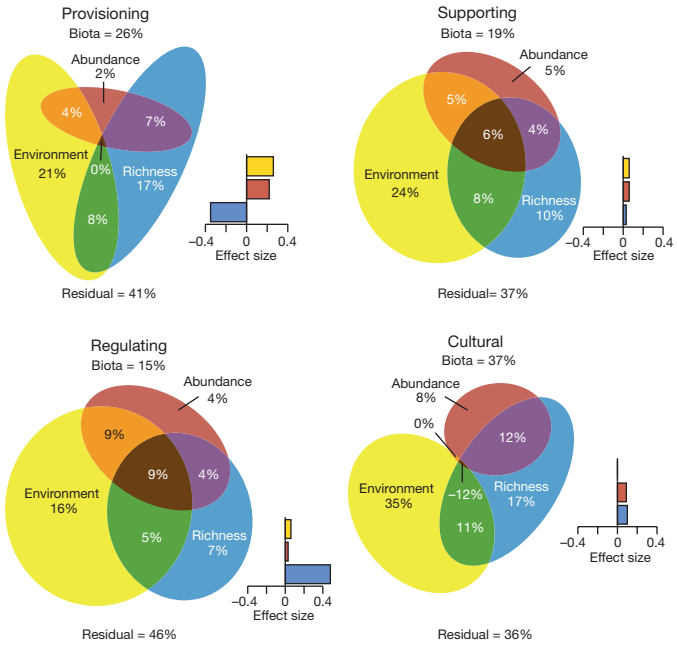


Figure 3 | Biotic versus abiotic drivers of ecosystem functioning. Variation partitioning for three predictor categories in our statistical models: environment, species richness and total abundance (details in Methods). Diagrams show the average across services within each ecosystem service type (detailed results are in Extended Data Figs 2 and 5). Shown are the unique variance explained by each predictor category, the shared variance between these categories (intersections of circles), and the variance not explained by the models (the residual). Biota refers to the total variance explained by species abundance and richness combined. Standardized effect sizes are shown as bar plots.

group generally complements the positive effects of other trophic groups on ecosystem services (see a comparison of multitrophic and unitrophic richness in Fig. 1). These contrasting effects caused multitrophic abundance to increase ecosystem multifunctionality only at low-to-moderate levels (Extended Data Fig. 1). Overall, our results underline the important role of species richness in driving the functioning of ecosystems[1,2,14–17,24,25], while also highlighting the often-overlooked effect of total biomass abundance on the supporting and provisioning services.

To test how generally applicable the trends in relationships between multitrophic richness and abundance were, we calculated multifunctionality metrics using all possible combinations of services. High multitrophic richness or abundance had increasingly positive effects as more services were considered and this effect was consistent across a wide range of levels of multifunctionality (Extended Data Fig. 3). To further explore this result, we calculated the similarities in the identities of the trophic groups driving a given pair of ecosystem services (the functional overlap, $\bar{o}$ (ref. 25)). On average, we found functional overlaps lower than 30% ($\bar{o} = 0.27 \pm 0.03$, mean ± s.e.m.), similar to results found for plant species in grassland experiments ($\bar{o} = 0.19$–0.49)[25]. This demonstrates low multitrophic redundancy and means that different services are supported by different trophic groups (Fig. 1 and Extended Data Fig. 1). We also found that different groups positively affected multifunctionality when it was calculated according to scenarios representing different land-use objectives (Extended Data Fig. 4). Finally, five of the nine trophic groups had the strongest net-positive effects on at least one ecosystem service (for example, primary producers on pest control, soil microbial decomposers on aggregate stability; Extended Data Fig. 1), with each group affecting at least one service. Collectively, these results show the low functional redundancy found between the multiple trophic groups studied, explaining why high multitrophic richness is needed to support high levels of ecological multifunctionality or to promote a larger number of ecosystem services.

The relationships between multitrophic richness, multitrophic abundance and ecosystem services were not always positive (Figs 1, 2 and Extended Data Fig. 1), consistent with previous studies[27,28]. Negative relationships might be explained by interference between species within a given trophic group[10] or by compositional shifts leading to declines in ecosystem functioning[17,20,29]. Despite these negative associations or harm to services, our results suggest that the most important trophic groups for maintenance of the services considered are aboveground herbivorous insects, primary producers and soil microbial decomposers. The richness or abundance of these trophic groups were most often correlated to ecosystem multifunctionality (43–72% of the 501 possible combinations between the services we measured), and had net-positive effects across all services (Fig. 2). These three groups also showed strong and frequent positive associations with the four main ecosystem service types (Fig. 1). These results agree with other studies that have identified plants and soil microorganisms as key drivers of ecosystem functioning[11,14,15,29], extending these findings to the richness and abundance of different trophic groups, including primary producers and consumers both above and below ground. The species richness of some of these functionally important trophic groups relate to whole-ecosystem diversity[3,30] and, thus, management strategies focused on them may foster synergies between biodiversity conservation and high multifunctionality levels.

The relative importance of both multitrophic richness and abundance compared to the environmental drivers of ecosystem functioning has been rarely studied outside of experiments or individual functions[7,8,11,15,21,24]. We therefore calculated the proportion of variance in ecosystem functioning that was explained by multitrophic richness, abundance and environmental (soil, topography and land-use) factors. Our models accounted for a large proportion (54–64%) of the variance in the provisioning, supporting, regulating and cultural ecosystem service types (Fig. 3 and Extended Data Fig. 5). Multitrophic richness and abundance explained at least as much of the variance in ecosystem functioning as abiotic conditions or land-use intensity did, and generally had stronger effects (Fig. 3; Extended Data Fig. 5). These results provide evidence that biodiversity is of comparative importance to environmental factors in driving ecosystem functioning. This is true

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

not only for individual functions in small-scale experiments[7,8], but also for multiple ecosystem services in realistic landscapes (see refs 15, 24).

Our study shows that the functional importance of biodiversity in real-world ecosystems has been greatly underestimated, as a result of focussing on individual trophic groups. We demonstrate here that the functional effects of multitrophic richness and abundance are as strong as, or even stronger than, those of the environment or land-use intensity. We identified primary producers, above-ground herbivores and soil decomposers as particularly important trophic groups for maintaining a functioning ecosystem. Our results suggest that it is important to preserve high levels of species richness, abundance or both within a wide range of taxa. This must include taxa often ignored by conservation efforts such as soil microbial decomposers[15], or those considered pests in agricultural systems such as herbivorous insects, if we are to promote high levels of the multiple ecosystem services upon which human well-being depends.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 March; accepted 7 July 2016.
Published online 17 August; corrected online 24 August 2016 (see full-text HTML version for details).

1. Cardinale, B. J. *et al.* Biodiversity loss and its impact on humanity. *Nature* **486,** 59–67 (2012).
2. Naeem, S., Duffy, J. E. & Zavaleta, E. The functions of biological diversity in an age of extinction. *Science* **336,** 1401–1406 (2012).
3. Allan, E. *et al.* Interannual variation in land-use intensity enhances grassland multidiversity. *Proc. Natl Acad. Sci. USA* **111,** 308–313 (2014).
4. Newbold, T. *et al.* Global effects of land use on local terrestrial biodiversity. *Nature* **520,** 45–50 (2015).
5. Naeem, S., Hahn, D. R. & Schuurman, G. Producer-decomposer co-dependency influences biodiversity effects. *Nature* **403,** 762–764 (2000).
6. Balvanera, P. *et al.* Linking biodiversity and ecosystem services: current uncertainties and the necessary next steps. *Bioscience* **64,** 49–57 (2014).
7. Hooper, D. U. *et al.* A global synthesis reveals biodiversity loss as a major driver of ecosystem change. *Nature* **486,** 105–108 (2012).
8. Tilman, D., Reich, P. B. & Isbell, F. Biodiversity impacts ecosystem productivity as much as resources, disturbance, or herbivory. *Proc. Natl Acad. Sci. USA* **109,** 10394–10397 (2012).
9. Petchey, O. L., McPhearson, P. T., Casey, T. M. & Morin, P. J. Environmental warming alters food-web structure and ecosystem function. *Nature* **402,** 69–72 (1999).
10. Duffy, J. E. *et al.* The functional role of biodiversity in ecosystems: incorporating trophic complexity. *Ecol. Lett.* **10,** 522–538 (2007).
11. Jing, X. *et al.* The links between ecosystem multifunctionality and above- and belowground biodiversity are mediated by climate. *Nat. Commun.* **6,** 8159 (2015).
12. Douglass, J. G., Duffy, J. E. & Bruno, J. F. Herbivore and predator diversity interactively affect ecosystem properties in an experimental marine community. *Ecol. Lett.* **11,** 598–608 (2008).
13. Deraison, H., Badenhausser, I., Loeuille, N., Scherber, C. & Gross, N. Functional trait diversity across trophic levels determines herbivore impact on plant community biomass. *Ecol. Lett.* **18,** 1346–1355 (2015).
14. Handa, I. T. *et al.* Consequences of biodiversity loss for litter decomposition across biomes. *Nature* **509,** 218–221 (2014).
15. Delgado-Baquerizo, M. *et al.* Microbial diversity drives multifunctionality in terrestrial ecosystems. *Nat. Commun.* **7,** 10541 (2016).
16. Garibaldi, L. A. *et al.* Wild pollinators enhance fruit set of crops regardless of honey bee abundance. *Science* **339,** 1608–1611 (2013).
17. McGrady-Steed, J., Harry, P. M. & Morin, P. J. Biodiversity regulates ecosystem predictability. *Nature* **390,** 162–165 (1997).
18. Grime, J. P. Benefits of plant diversity to ecosystems: immediate, filter and founder effects. *J. Ecol.* **86,** 902–910 (1998).
19. Hillebrand, H., Bennett, D. M. & Cadotte, M. W. Consequences of dominance: a review of evenness effects on local and regional ecosystem processes. *Ecology* **89,** 1510–1520 (2008).
20. Soliveres, S. *et al.* Locally rare species influence grassland ecosystem multifunctionality. *Phil. Trans. R. Soc. B* **371,** 20150269 (2016).
21. Grace, J. B. *et al.* Integrative modelling reveals mechanisms linking productivity and plant species richness. *Nature* **529,** 390–393 (2016).
22. Byrnes, J. E. K. *et al.* Investigating the relationship between biodiversity and ecosystem multifunctionality: challenges and solutions. *Methods Ecol. Evol.* **5,** 111–124 (2014).
23. Millennium Ecosystem Assessment. *Ecosystems and Human Well-Being* 1–86 World Resources Institute (2005).
24. Maestre, F. T. *et al.* Plant species richness and ecosystem multifunctionality in global drylands. *Science* **335,** 214–218 (2012).
25. Hector, A. & Bagchi, R. Biodiversity and ecosystem multifunctionality. *Nature* **448,** 188–190 (2007).
26. Bruno, J. F., Boyer, K. E., Duffy, J. E. & Lee, S. C. Relative and interactive effects of plant and grazer richness in a benthic marine community. *Ecology* **89,** 2518–2528 (2008).
27. Balvanera, P. *et al.* Quantifying the evidence for biodiversity effects on ecosystem functioning and services. *Ecol. Lett.* **9,** 1146–1156 (2006).
28. Lefcheck, J. S. *et al.* Biodiversity enhances ecosystem multifunctionality across trophic levels and habitats. *Nat. Commun.* **6,** 6936 (2015).
29. Naeem, S., Thompson, L. J., Lawler, S. P., Lawton, J. H. & Woodfin, R. M. Declining biodiversity can alter the performance of ecosystems. *Nature* **368,** 734–737 (1994).
30. Manning, P. *et al.* Grassland management intensification weakens the associations among the diversities of multiple plant and animal taxa. *Ecology* **96,** 1492–1501 (2015).

Acknowledgements We thank B. Schmid, F. T. Maestre and S. Kéfi for comments that helped improve this manuscript. W. Ulrich and N. J. Gotelli provided statistical advice. We thank the people who maintain the Biodiversity Exploratories program: A. Hemp, K. Wells, S. Gockel, K. Wiesner and M. Gorke (local management team); S. Pfeiffer and C. Fischer (central office), B. König-Ries and M. Owonibi (central database management); and E. Linsenmair, D. Hessenmöller, J. Nieschulze, E.-D. Schulze and the late E. Kalko for their role in setting up the project. This work was funded by the Deutsche Forschungsgemeinschaft Priority Program 1374 'Infrastructure-Biodiversity Exploratories'. Fieldwork permits were given by the responsible state environmental offices of Baden-Württemberg, Thüringen and Brandenburg (according to §72 BbgNatSchG). Figure icons were created by R. D. Manzanedo.

Author Contributions S.S. and E.A. conceived the idea of this study. M.F. initiated the Biodiversity Exploratories project aimed at measuring multiple diversities and functions in the field sites. All authors but S.S., E.A. and F.V.D.P. contributed data. S.S. and F.V.D.P. performed the analyses. S.S. and S.C.R. performed the literature search. S.S. wrote the first draft of the manuscript and all the authors (especially E.A., P.M., F.D.V.P., M.M.G. and D.P.) contributed substantially to the revisions.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.S. (santiago.soliveres@ips.unibe.ch).

Reviewer Information *Nature* thanks Y. Hautier, F. Isbell and the other anonymous reviewer(s) for their contribution to the peer review of this work.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Study sites. We selected a total of 150 grassland sites (50 m × 50 m) in three regions of Germany (50 sites per region) to cover a gradient of land-use intensities, characterized by contrasting grazing, fertilization and mowing levels (www.biodiversity-exploratories.de, ref. 31). The regions in the south-west (Schwäbische Alb) and the north-east (Schorfheide-Chorin) are UNESCO Biosphere Reserves, whereas the central region is in and around the Hainich National Park. The three regions differ substantially in geology, climate and topography[31], covering a range of ~3 °C in mean annual temperature and 500 mm in annual precipitation. Plots in each region cover the range of land-use intensities typical for Central European grasslands. We obtained information on land use via questionnaires sent to land owners, asking about the number and type of livestock (converted to livestock units) and the duration of grazing in each plot, the fertilization (from which we calculated the amount of nitrogen added), and the mowing (number of cuts per year[31,32]). We used this information to calculate three standardized indices summarizing grazing, fertilization and mowing intensity (see ref. 32 for full methodological details).

Diversity measures. At each site, we measured the species richness and abundance of nine functional groups using standard methodologies (Extended Data Table 2). In total we observed about 4,600 taxa on the 150 grasslands studied. The nine trophic groups were: primary producers (vascular plants and bryophytes), below-ground herbivores (herbivorous insect larvae sampled in the soil), below-ground predators (carnivorous insect larvae sampled in the soil), detritivores (insects and Diplopoda feeding on leaf litter and other detritus), soil microbial decomposers (soil bacteria), above-ground herbivores (insects feeding solely on above-ground plant material), above-ground predators (carnivorous insects, spiders and Chilopoda), plant symbionts (arbuscular mycorrhizal fungi), and bacteria-feeding protists (heterotrophic flagellates and ciliates). Lichens and omnivores were not considered in our analyses as they were too rare. We directly measured species richness for most groups, but richness was quantified as family richness for below-ground insects and soil bacteria and as the number of operational taxonomic units (OTUs) for the mycorrhizae and protists. The abundance of each trophic group was also measured using different methods: number of individuals for arthropods, amount of cover for vascular plants and bryophytes, and relative proportion of sequence reads assigned to each family or OTU for protists, soil bacteria and mycorrhiza. To avoid multicollinearity, we did not include the abundances of protists or detritivores as they were highly correlated (Spearman's $\rho > 0.6$) with richness (for more details see Extended Data Table 3).

We also measured the abundance and richness of foliar fungal pathogens, pollinators and birds; however, to include a broader range of ecosystem services in our analyses, we treated these groups as proxies of ecosystem services. Total pollinator abundance and the inverse of pathogen abundance were treated as proxies of regulating services (pollination and disease regulation), and we used bird-species richness as a measure of a cultural ecosystem service. Lepidoptera behave as herbivores during juvenile stages and as pollinators when adults. To avoid accounting for them twice, we assigned them to only one trophic group (pollinators), as the data were counts of the adult butterflies, not the caterpillars.

Ecosystem functioning measures. At each site, we measured 14 different ecosystem variables (both functions and service proxies; Extended Data Table 2) and classified them into four types of services following the Millennium Ecosystem Assessment[23]. These 14 ecosystem services were: i) supporting services related to nutrient capture and cycling (root biomass, root decomposition rates, potential nitrification (based on urease activity in soil samples), phosphorus retention (calculated as a ratio between shoot and microbial phosphorus stock and soil extractable phosphorus), arbuscular mycorrhizal fungal root colonization (measured as hyphal length), soil aggregate stability (proportion of water-stable soil aggregates)); ii) provisioning services related to agricultural value (forage production (above-ground plant biomass) and forage quality (based on crude protein and relative forage value); iii) regulating services for neighbouring crop production or climate regulation (that is, regulating services: resistance to plant pathogens, pest control, pollinator abundance and soil organic carbon); or iv) cultural services linked to recreation (bird diversity and flower cover). Because the values for trophic groups and ecosystem functions varied widely, we standardized all variables to a common scale ranging from 0 to 1 according to the following formula: $\mathrm{STD} = (X - X_{\min})/(X_{\max} - X_{\min})$; where STD is the standardized variable and X, $X_{\min}$ and $X_{\max}$ are the target variable, and its minimum and maximum value across all sites, respectively. This made slope estimates for different predictors comparable.

We calculated ecosystem multifunctionality metrics from the 14 services as the percentage of measured services (measured services only to correct for the fact that some services that had not been measured in all sites) that exceeded a given threshold of their maximum observed level across all study sites. To reduce the influence of outliers we calculated the maximum observed level as the average of the top five sites[22,33]. Given that any threshold is likely to be arbitrary, the use of multiple thresholds is recommended to better understand the role that biodiversity plays in affecting ecosystem multifunctionality and to understand trade-offs between functions of interest[22]. Therefore, we used four different thresholds (25%, 50%, 75% and 90%) to represent a wide spectrum in the analyses performed (Extended Data Figs 1–3). As an alternative approach we also calculated multifunctionality scenarios, weighting the services differently according to the different potential views of stakeholders (that is, stakeholders willing only to promote provisioning services versus those trying to maximize cultural and recreation services or the sustainability of soils and crops; Extended Data Fig. 4)[34].

Effects of multitrophic richness and abundance on grassland ecosystem services and multifunctionality. We used linear models to evaluate the relationships between species richness and abundance in the nine trophic groups and each of the 14 individual ecosystem services, along with the different multifunctionality metrics (four thresholds and the metrics were obtained by weighting each ecosystem service according to different potential stakeholders' needs; for example, only provision, sustainable soils and crops, or cultural scenarios, see ref. 34). In all cases, we used a Gaussian error distribution as the errors of our response variables were normally distributed. We report the effects of the different trophic groups on the different functions as slopes from the multiple regression model; these are corrected for the effects of all other variables in the model. Since our main focus was on calculating the independent effects of the richness and abundance of the different trophic groups, we corrected them for co-varying factors. Thus, we calculated residuals for all our variables (both biotic predictors and functioning measures) from linear models including region, land-use intensity (standardized measures of mowing, grazing and fertilization intensity) and other important environmental factors (soil type and depth, pH, a topographic wetness index based on position within the slope and orientation, and elevation). As an alternative to using residuals, we also fitted models with all the environmental and land-use predictors (standardized to give comparable coefficients) alongside the diversity and abundance measures. These approaches gave very similar results (Extended Data Fig. 2). Standardized coefficients of the functional effects of richness were very similar, whether or not abundance was included ($\rho = 0.80$, $P < 0.0001$, $N = 162$; data not shown). We also fitted models with the abundance and richness of only one individual trophic group to compare the results of the best individual trophic group with the multitrophic analyses (Extended Data Fig. 1). Finally, we fitted models with only richness of vascular plant species as a predictor. The latter is the most common measure of biodiversity[7,8,21,24,25,35–38] and we used it to compare our results with those found in previous studies on biodiversity–ecosystem functioning relationships.

We performed model simplification using the stepAIC function in R, and further simplified the minimal models produced using AIC by removing all terms that were not significant according to F-ratio tests (Extended Data Table 4). Results using alternative approaches for model selection are presented in Extended Data Fig. 6. We did not fit interactions between the richness and abundance of different trophic groups, or between those and environmental factors, as this would require a large number of coefficients, would be difficult to interpret and would require an even larger data set than ours (see ref. 20 for a study evaluating the interaction between land-use and diversity). We did not find evidence of nonlinear relationships between our predictors and the ecosystem services measured when checking all bivariate relationships; thus we did not include nonlinear terms in the models to keep them simple.

Not all trophic groups or ecosystem services were measured on all sites; thus different services were analysed using different sized data sets (N ranged between 111 and 54, depending on the service). The different sampling sizes used were not related to the number of trophic groups included in the most parsimonious model (Spearman's rank correlation coefficient $\rho = 0.32$), the increase in variance explained by vascular plant species richness ($\rho = -0.21$) or the net effect of richness or abundance ($\rho = -0.01$ or 0.05, respectively; $N = 14$ and $P > 0.25$ in all cases). Thus, fitting models differing in sample size for different services did not affect our results.

The inclusion of many predictors in statistical models increases the chance of type I error (false positives). To account for this we used a Bernoulli process to detect false discovery rates, where the probability (P) of finding a given number of significant predictors (K) just by chance is a proportion of the total number of predictors tested ($N = 16$ in our case: the abundance and richness of 7 and 9 trophic groups, respectively) and the P value considered significant ($\alpha = 0.05$ in our case)[39,40]. The probability of finding three significant predictors on average, as we did, is therefore, $P = [16!/(16-3)!3!] \times 0.053(1-0.05)^{(16-3)} = 0.0359$, indicating

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

that the effects we found are very unlikely to be spurious. The probability of false discovery rates when considering all models and predictors fit (14 ecosystem services × 16 richness and abundance metrics) and the ones that were significant amongst them (52: 25 significant abundance predictors and 27 significant richness predictors) was even lower ($P < 0.0001$). All analyses were performed using R version 3.0.2 (ref. 41).

Net functional effects of the different trophic groups across ecosystem service types. We calculated the net effect of each trophic group on each ecosystem service type (provisioning, supporting, regulating and cultural) by fitting all services belonging to these types into a single model. To do so, we added two extra predictors to our models: 'service identity' as a fixed factor, to account for differences between individual services, and 'site' as a random factor, to account for correlations between services, abundance and richness values measured on the same site. Since we were interested in the net effects of each group across all services, we did not fit interactions between our multitrophic predictors and service identity. The net effect across all services was analysed using the same approach, while fitting a single model for the 14 ecosystem services at the same time. This approach corrects for the fact that the individual service models vary in their explanatory power and in the predictor variables included. Fitting all services into a single model allows us to obtain a robust estimate of the net functional effect (the standardized coefficient from the model) of the abundance and richness of each trophic group on the service type of each ecosystem and on ecosystem multifunctionality, together with an estimate of its error. If the standardized coefficient was positive, we interpreted it as a net overall positive effect of either richness or abundance across all services, or on a given service type (Figs 1 and 2). In all cases, we used standardized coefficients of the most parsimonious models after model reduction. However, our results remained when using other approaches that account for differences in model fit, such as multi-model averaging coefficients (coefficients were weighted according to the AIC weight of the models in which each predictor is included) or when weighting the standardized coefficient for each ecosystem service by the adjusted R^2 of each model (which should also be comparable across models with different response variables; Extended Data Fig. 6).

Variance partitioning analyses. Variance partitioning analyses (also known as commonality analyses) were performed with standard techniques[42,43] based on the comparison of variance explained by models including every possible combination of variables. Variables were organized by environment (study region, soil type, pH, topographic wetness index, grazing and fertilization, with the remaining environmental predictors removed to prevent multicollinearity; shown in Extended Data Table 3), species richness (standardized species richness of the nine trophic groups) and abundance (standardized abundance of those trophic groups in which abundance and richness were not strongly correlated ($\rho < 0.6$; shown in Extended Data Table 3)). Thus, we fitted a series of seven models for each service and multifunctionality metric (at the 25%, 50%, 75% and 90% thresholds) to extract the unique and shared variance for each combination of variables (environment only, richness only, abundance only, environment + abundance, environment + richness, richness + abundance, and all predictors together). Variance-partitioning analyses were performed with the full models (without model simplification) to allow us to compare between different services. As a consequence, we used R^2 rather than the adjusted R^2 because, owing to the large number of predictors, some adjusted R^2 values were negative, complicating the extraction of unique variance explained by each predictor. Venn diagrams were drawn using Euler APE for Windows[44].

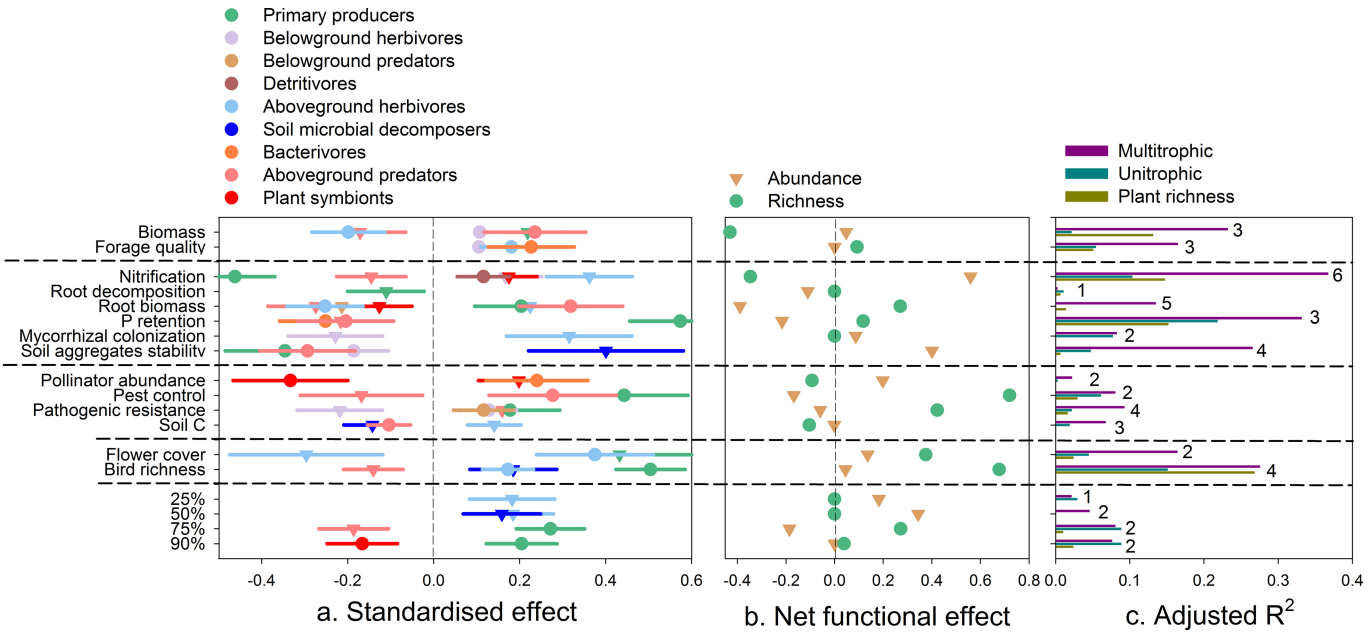
To compare the effect size among richness, abundance and environment on the different ecosystem services and multifunctionality metrics, we summed the standardized coefficients of all predictors from each component (the abundance of five trophic groups (abundance), the richness of nine trophic groups (richness), and pH, fertilization, grazing, and topographic wetness index (environment)). We excluded study region and soil type when summing effects, as these were categorical predictors and their coefficients were not straightforward to interpret. We performed these calculations for each of the 14 ecosystem services and 4 multifunctionality metrics in isolation (Extended Data Fig. 4), and for each ecosystem service type (Fig. 2) by using models containing all the ecosystem services belonging to each type into a single model (again, adding 'service identity' and 'site' as fixed and random predictors, respectively).

Analysing every possible combination of ecosystem services. Studies on multifunctionality are difficult to compare as they include different measures of ecosystem functioning. To allow us to generalize our results and to test whether multitrophic richness and abundance are more important in supporting higher numbers of services simultaneously, we also calculated multifunctionality indices using every possible combination of the services we measured. We did this after removing those services with more than 20 missing sites, leaving a total of 9 services (501 combinations) as response variables. We calculated multifunctionality at the 25%, 50%, 75% and 90% thresholds for all these combinations (Extended Data Fig. 3). We also tested the sensitivity of our analyses to missing data by repeating our analyses for every possible combination of 1–13 of the 14 measured services (16,368 combinations; results for multifunctionality calculated with all 14 ecosystem services are presented in Extended Data Figs 1, 2). To allow the comparison of models with different services, data gaps were filled with the average value of a given service in a given region, which is a conservative approach. In both cases (combinations of 1–13 or 1–9 functions), the most parsimonious models possible were selected on the basis of their AIC. This avoids inflated type I error, caused by fitting a large number of models, as model selection was not based on P values. Results using 9 or 14 functions were qualitatively the same and therefore only the former are shown here.

Review of multitrophic manipulative approaches. Manipulative experiments including as many groups and services as we considered in this study do not yet exist. However, we compared our correlational results with available evidence from experiments manipulating the diversity of more than one trophic group. To do this we performed a bibliographic research in the Web of Knowledge and in Google Scholar using all combinations of the terms 'multitrophic' or 'trophic groups' + 'functioning' or 'multifunctionality' or 'biomass' or 'ecosystem services' or 'diversity'. We also screened references within available reviews on multitrophic diversity–ecosystem functioning relationships[10,27,45]. Of the papers found, we selected those which fulfilled the following criteria: i) it was a manipulative study, ii) it included a range in species richness (not only presence or absence) of, at least, two different trophic groups and iii) it provided enough information to calculate the increase in variance explained by the addition of a second trophic group. Only four studies, including seven ecosystem functions, fulfilled these criteria (Extended Data Table 1). Some of these manipulative studies did not include plants, so we calculated the percentage increase in variance seen when comparing a model with the trophic group that had the strongest explanatory power in models containing two trophic groups. When the same function was measured across several studies (that is, biomass), we calculated the average increase in variance explained for this variable when another trophic level was added. These results were used to qualitatively compare the limited evidence from multitrophic manipulations with our results.

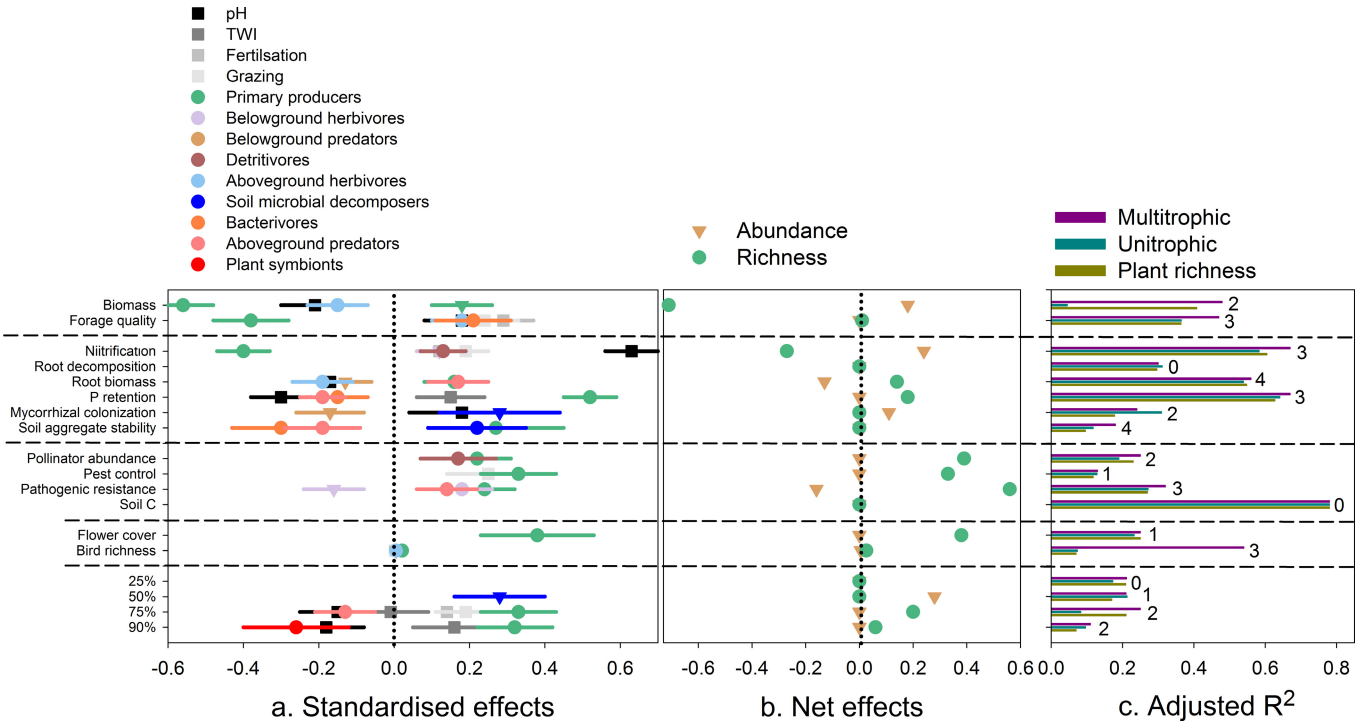
31. Fischer, M. *et al.* Implementing large-scale and long-term functional biodiversity research: The Biodiversity Exploratories. *Basic Appl. Ecol.* **6,** 473–485 (2010).
32. Blüthgen, N. *et al.* A quantitative index of land-use intensity in grasslands: integrating mowing, grazing and fertilization. *Basic Appl. Ecol.* **13,** 207–220 (2012).
33. Zavaleta, E. S., Pasari, J. R., Hulvey, K. B. & Tilman, G. D. Sustaining multiple ecosystem functions in grassland communities requires higher biodiversity. *Proc. Natl Acad. Sci. USA* **107,** 1443–1446 (2010).
34. Allan, E. *et al.* Land use intensification alters ecosystem multifunctionality via loss of biodiversity and changes to functional composition. *Ecol. Lett.* **18,** 834–843 (2015).
35. Hautier, Y. *et al.* Eutrophication weakens stabilizing effects of diversity in natural grasslands. *Nature* **508,** 521–525 (2014).
36. Gamfeldt, L., Hillebrand, H. & Jonsson, P. R. Multiple functions increase the importance of biodiversity for overall ecosystem functioning. *Ecology* **89,** 1223–1231 (2008).
37. Gamfeldt, L. *et al.* Higher levels of multiple ecosystem services are found in forests with more tree species. *Nat. Commun.* **4,** 1340 (2013).
38. Cardinale, B. J. *et al.* Effects of biodiversity on the functioning of trophic groups and ecosystems. *Nature* **443,** 989–992 (2006).
39. Moran, M. D. Arguments for rejecting the sequential Bonferroni in ecological studies. *Oikos* **100,** 403–405 (2003).
40. Tylianakis, J. M. *et al.* Resource heterogeneity moderates the biodiversity-function relationship in real world ecosystems. *PLoS Biol.* **6,** e122 (2008).
41. R Development Core Team. *R: A Language And Environment For Statistical Computing* (R Foundation For Statistical Computing, 2014).
42. Borcard, D., Legendre, P. & Drapeau, P. Partialling out the spatial component of ecological variation. *Ecology* **73,** 1045–1055 (1992).
43. Peres-Neto, P. R., Legendre, P., Dray, S. & Borcard, D. Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology* **87,** 2614–2625 (2006).
44. Micallef, L. & Rodgers, P. eulerAPE: drawing area-proportional 3-Venn diagrams using ellipses. *PLoS One* **9,** e101717 (2014).
45. Worm, B. & Duffy, J. E. Biodiversity, productivity and stability in real food webs. *Trends Ecol. Evol.* **18,** 628–632 (2003).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



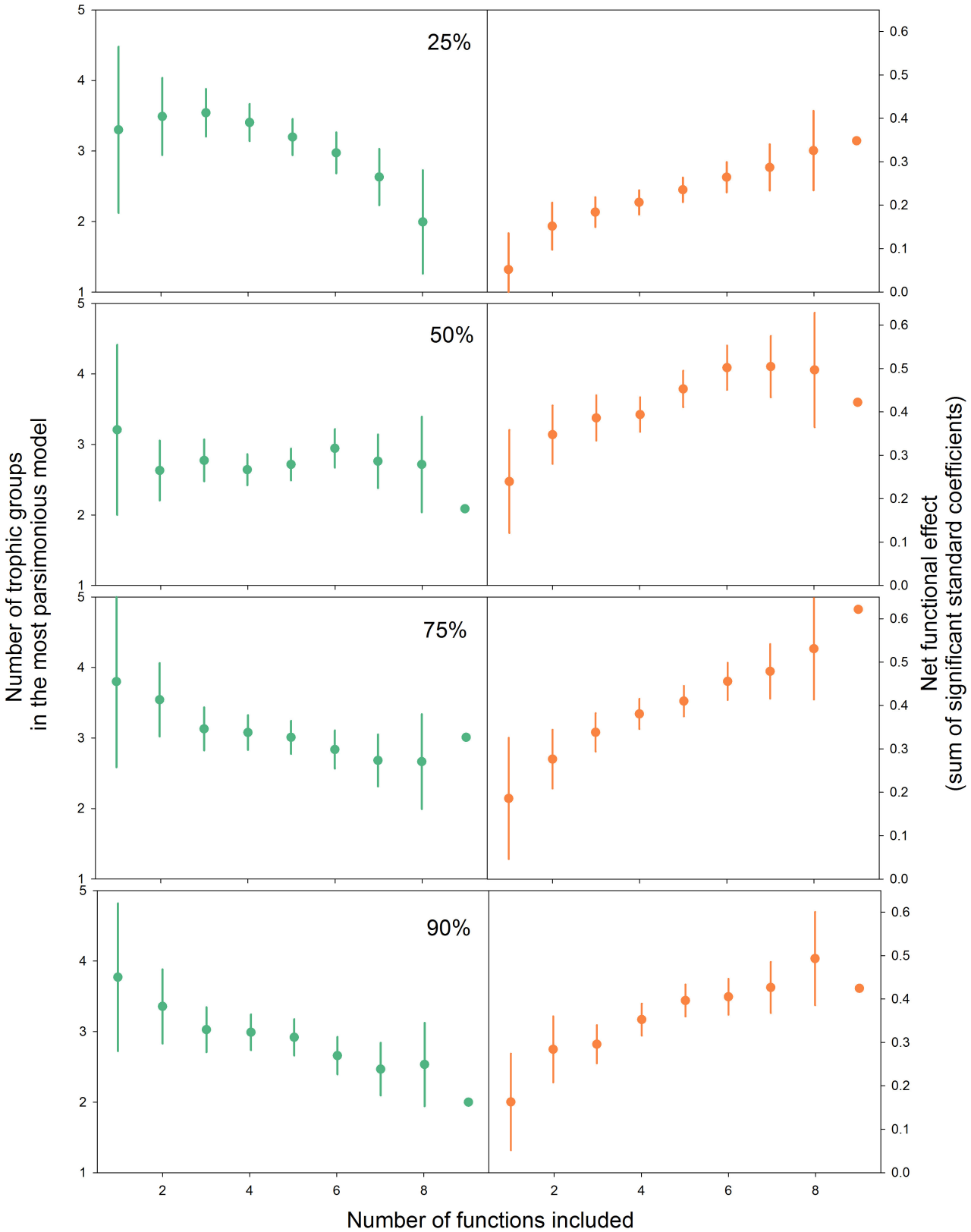
Extended Data Figure 1 | Functional effects of multitrophic richness and abundance on 14 grassland ecosystem services. a, Standardized coefficients (mean ± s.e.m.) of the abundances (triangles) and richness (circles) of those trophic groups that significantly affect a given function are shown. **b**, The net effect (that is, the sum of significant standardized effects). **c**, Difference in adjusted R^2 between the final multitrophic models and those models using the abundance and richness of the best performing individual trophic group (unitrophic) or plant species richness (plant richness). Ecosystem services are organized by the main four types of services they associate with (provisioning, supporting, regulating and cultural). The number of trophic groups included in the most parsimonious model is given next to their adjusted R^2. Multifunctionality results at 25%, 50%, 75% and 90% thresholds are also shown (see Methods).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



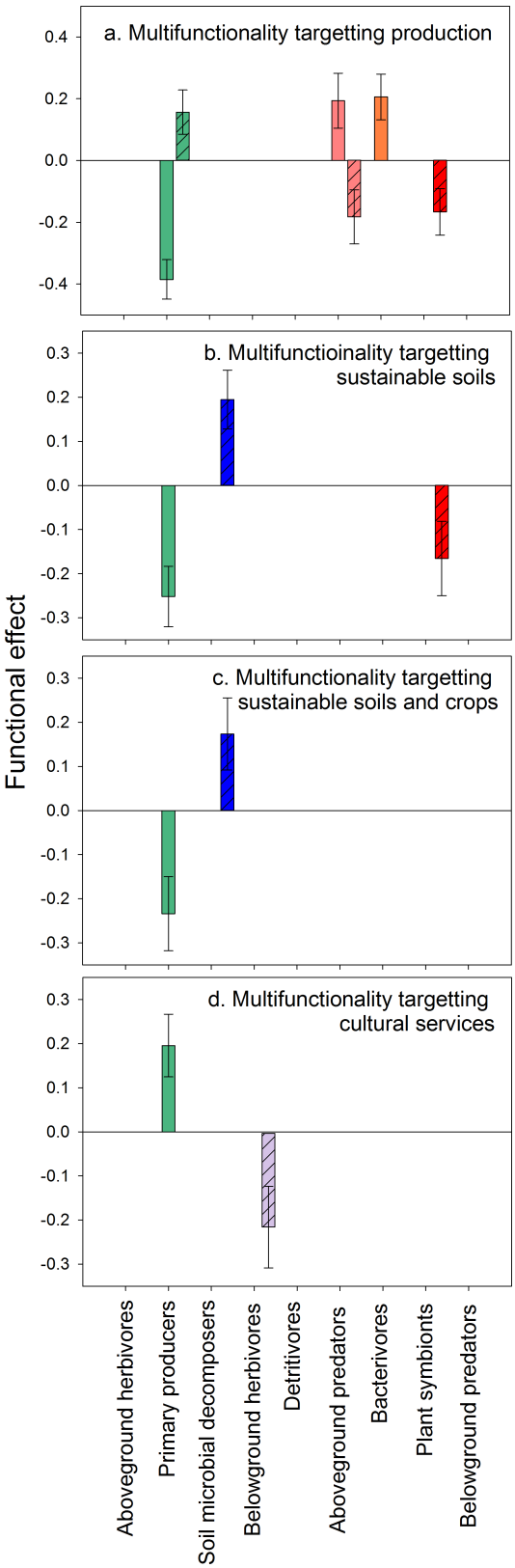
Extended Data Figure 2 | Functional effects of environmental factors and multitrophic richness and abundance on 14 grassland ecosystem functions. a, Standardized slope estimates (mean ± s.e.m.) for each significant predictor are shown, with the exception of study region and soil type, which were retained in all models. **b**, Net effect (sum of significant standardized effects) for multitrophic richness and abundance. **c**, The total amount of variance explained by either environmental + plant species richness, environmental + the abundance and richness of the best individual trophic predictor, or by environmental + multitrophic diversity and abundance are shown for each function (adjusted R^2, to control for the high number of predictors included). The number of trophic groups included in the best models (2.15 ± 1.2 across functions, and 1.94 ± 1.2 across functions and multifunctionality indices) is given next to the adjusted R^2 value. The increase in the adjusted R^2 values in models with plant-species-richness averaged 0.07 ± 0.12 (across functions) and 0.06 ± 0.11 (across functions and multifunctionality indices). Ecosystem services are organized by the main four types of services they associate with (top–bottom: provisioning, supporting, regulating and cultural). TWI, topographic wetness index, based on the aspect and position in the slope, and the inclination of the slope. Multicollinearity between the predictors introduced is unlikely (Extended Data Table 3).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



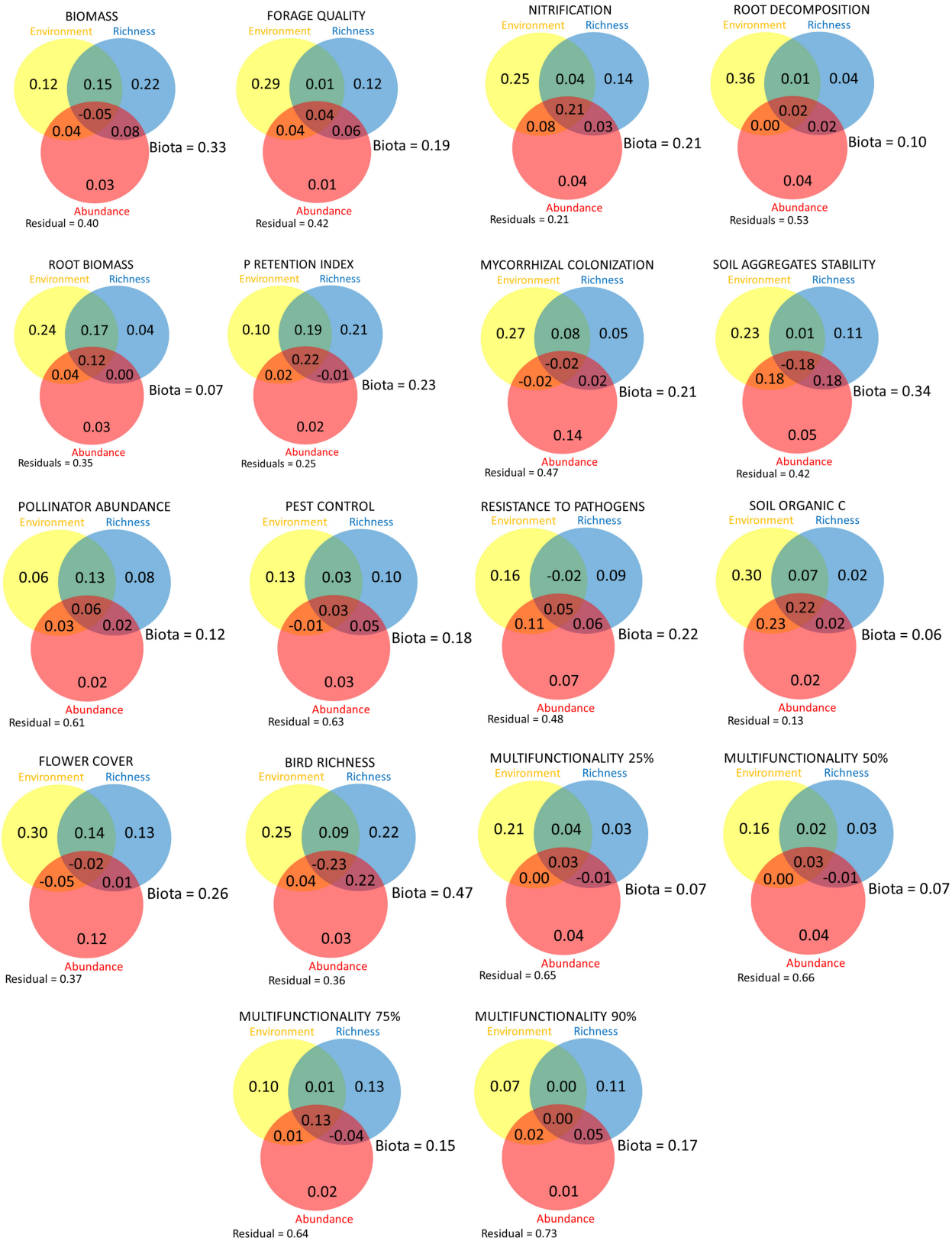
Extended Data Figure 3 | Number of trophic groups necessary to predict multifunctionality measures calculated with all possible combinations of 1–9 services, and their net effects. The number of predictors selected in the best models (left) and their overall effects (sum of standardized coefficients; right) across all possible combinations of 1–9 services ($N = 501$) are shown. Error bars show the 95% confidence intervals, estimated for all possible combinations of n (1 to 9) functions in both cases. Only the 9 services with fewer than 20 data gaps were considered in these analyses (see details in Methods). Multifunctionality for these combinations was calculated at the 25% (upper panel), 50%, 75% and 90% (bottom panel) thresholds. Services removed were flower cover, arbuscular mycorrhizal colonization, soil aggregate stability, phosphorous retention index and pest control.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



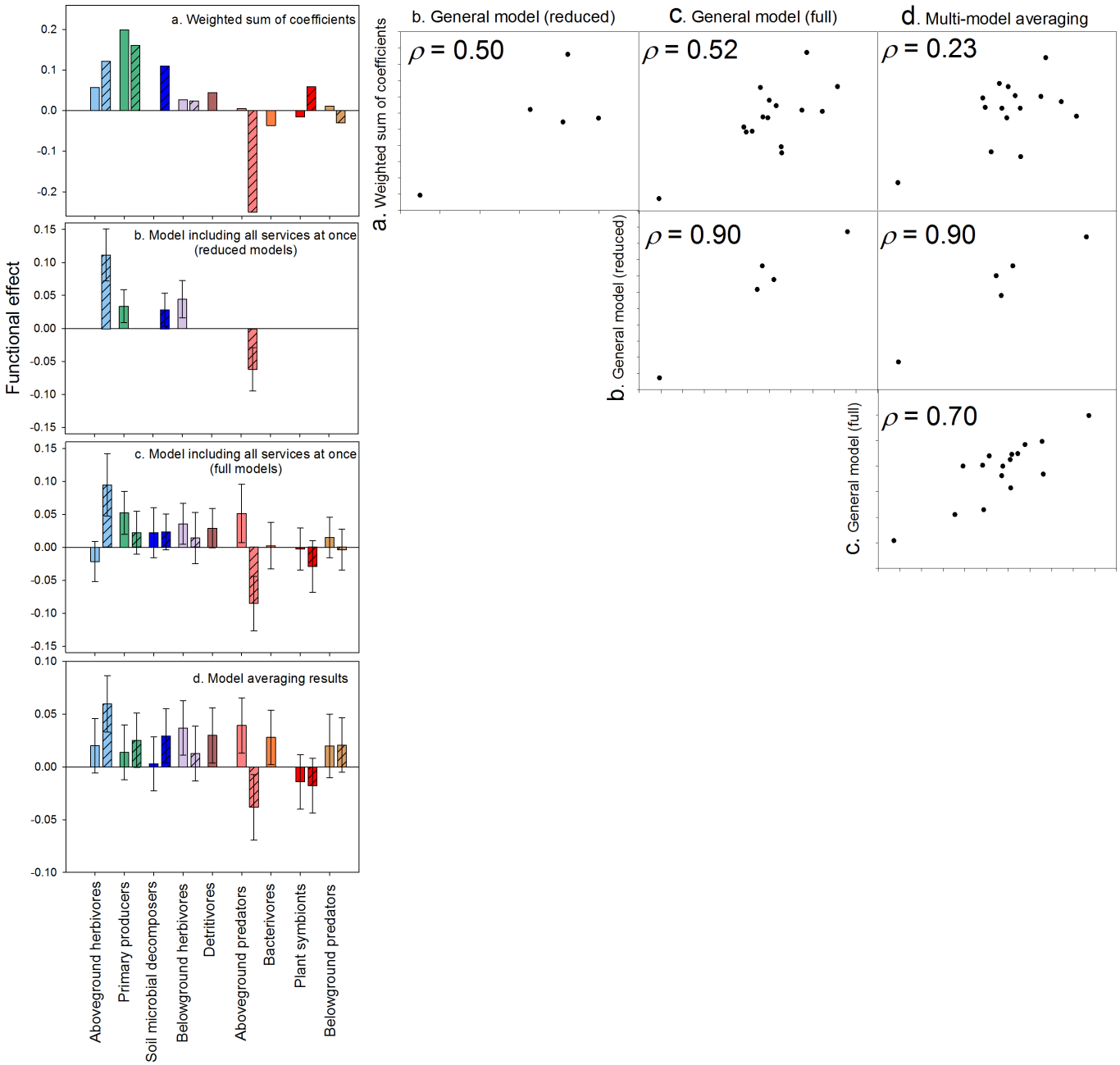
Extended Data Figure 4 | Functional effect of the different trophic groups on contrasting multifunctionality scenarios. Overall functional effects (significant standardized coefficients; mean ± s.e.m.) from the most parsimonious model) of the richness (open bars) and abundance (hatched bars) of each group are shown according to ref. 34.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 5 | Functional importance of species richness and abundance compared to environmental drivers. Venn diagrams showing the variance partition for the four components of our statistical models (environment: climate, soil and land-use intensity; species richness of the nine trophic groups, abundance of primary producers, above- and below-ground predators, below-ground herbivores and soil microbial decomposers). The variance not explained by the model (the residual) is also shown. The variance explained by richness, abundance and their overlap is summed up as Biota. Each panel represents an individual function or multifunctionality metric.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 6 | Functional effect of the different trophic groups. Overall functional effects (mean ± s.e.m. of the standardized slopes obtained from the model; with the exception of **a**, where error could not be estimated) of the richness (open bars) and abundance (hatched bars) of each group. **a**, The values were calculated after weighting each standardized coefficient (those in Extended Data Fig. 1) by the adjusted R^2 of the model to account for differences in model performance. **b**, **c**, The values were calculated as the standardized coefficients in a general model fitted to all services at once, including 'service identity' as an extra predictor and 'plot' as random factor to control for pseudo-replication (reduced models (**b**); the ones presented in the main text), or full models (**c**) and, **d**, calculated as multi-model average parameters from a model fitted to all services at once. Correlations (Spearman's rank correlation coefficients) between the different approaches are given.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 1 | Re-analysis of manipulative multitrophic studies

Study ID	Reference	System	# trophic groups manipulated	Approach	Response variable	Variance explained best group	Variance explained two trophic groups	% increase with multitrophic approach	comment
1	Douglass et al. 2008 Ecolett	aquatic	2	mesocosm	grazer abundance	55.0	61.0	10.9	based on ω^2
1	Douglass et al. 2008 Ecolett	aquatic	2	mesocosm	predator abundance	5.0	7.0	40.0	based on ω^2
2	Bruno et al. 2008. Ecology	aquatic	2	mesocosm	autotroph biomass	19.5	42.7	119.0	based on F
3	Naeem et al. 2000. Nature	terrestrial	2	microcosm	autotroph biomass	13.6	23.5	72.8	based on F
3	Naeem et al. 2000. Nature	terrestrial	2	microcosm	detritivore biomass	5.6	11.3	101.8	based on F
3	Naeem et al. 2000. Nature	terrestrial	2	microcosm	# C sources used	7.2	13.5	87.5	based on F
4	Handa et al. 2014 Nature	both	2	field.expt	Litter C loss	5.8	6.6	13.8	based on %SS

For each study, an ID number and full reference are given. The system in which each study was performed (aquatic or terrestrial), the number of trophic groups manipulated and the approach used (controlled mesocosms or field studies) are provided. The ecosystem functions ('response variable') measured within each study were grouped in biomass production (the first five rows), nutrient cycling (sixth row) and decomposition (seventh row). Variance explained (according to the statistic mentioned in comments; ω^2 = proportion of variance explained according to the authors; F = Fisher's F, SS = sum of squares) for the single trophic group with the most explanatory power, and the difference between the variance explained by this group and the inclusion of a second group are given (grey column).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 2 | Details of the sampling procedure for each trophic group and function

Trophic groups			
Trophic group	**Subgroup**	**Sampling method**	**Author**
Primary producers	Plants, bryophytes	Measurement of % cover in a 4×4 m subplot, done in 2009	Boch, Heinze, Hölzel, Klaus, Kleinebecker, Müller, Prati, Socher, Fischer
Aboveground herbivores	Herbivorous insects	Sweep netting (Hemiptera: Heteroptera/Auchenorrhyncha, Hymenoptera, Neuroptera and Orthoptera). Transects of 150m with 60 double sweeps, done twice per plot in 2008-2010.	Lange, Pašalić, Türke, Gossner, Weisser
Aboveground predators	Carnivorous insects	Sweep netting (Hemiptera: Heteroptera/Auchenorrhyncha, Hymenoptera, Neuroptera and Orthoptera). Transects of 150m with 60 double sweeps done twice per plot in 2008-2010.	Lange, Pašalić, Türke, Gossner, Weisser
	Spiders	Sweep netting. Transects of 150m with 60 double sweeps, done twice per plot in 2008-2010.	Lange, Pašalić, Türke, Gossner, Weisser
	Chilopoda	Kempson extraction from one soil core of 20 ×5 cm per plot, done in 2008	Birkhofer, Diekötter, Wolters
Detritivores	Annelids	Hand sorting from two soil cores of 20 ×10 cm per plot, done in 2008	Birkhofer, Diekötter, Wolters
	Diplopoda	Kempson extraction from one soil core of 20 ×5 cm per plot, done in 2008	Birkhofer, Diekötter, Wolters
	Detritivorous insects	Sweep netting (Hemiptera: Heteroptera/Auchenorrhyncha, Hymenoptera, Neuroptera and Orthoptera). Transects of 150m with 60 double sweeps, done twice per plot in 2008-2010.	Lange, Pašalić, Türke, Gossner, Weisser
Microbial decomposers	Soil bacteria	cDNA amplicon sequencing of partial (V3) 16S rRNA gene transcripts, done in 2011	Baumgartner, Sikorski, Overmann
Bacterivores	Bacterivorous protists	18S rDNA gene PCR and amplicon sequencing (454) filtering for rhizarians, alveolates, stramenopiles and opisthokonts, done in 2011	Venter, Arndt
Symbionts	Arbuscular mycorrhizal fungi	Pyrotag sequencing of the NS31 - AM1 fragment of the 18S rDNA genes, done in 2011	Klemmer, Wubet, Buscot
Belowground herbivores	Insect larvae	Extracted from a heat/moisture gradient in one soil core of 20 x 5 cm per site, done in 2011 over a period of eight days.	Sonnemann, Wurst
Belowground predators	Insect larvae	Extracted from a heat/moisture gradient in one soil core of 20 x 5 cm per site, done in 2011 over 8 days.	Sonnemann, Wurst
Functions			
Function		**Sampling method**	**Author**
Aboveground plant biomass		Harvested in four 0.5 m × 0.5 m quadrats per plot, done in May-June in 2008-2012.	Schmitt, Prati, Fischer, Klaus, Kleinebecker, Hölzel
Belowground plant biomass		Measured in 14 soil cores (0-10 cm). Fine roots were sorted and weighted after drying in the oven, done in samples collected in May 2011	Solly, Schöning, Schrumpf
Root decomposition rate		Measured as the mass loss from root litter bags after 6 months, from October 2011 to April 2012.	Solly, Schöning, Schrumpf
Potential nitrification		10 mM ammonium sulphate solution was added as substrate to 2.5g of soil composite samples (i.e. the same samples as for soil carbon; see below).	Stempfhuber, Schloter
Phosphorus uptake and retention		Proportion of P in plants and microbes (shoot P stock + microbial P stock) / (shoot P stock + microbial P stock + soil extractable P [$NaHCO_3$]).	Alt, Sorkau, Oelman, Wilcke, Klaus, Kleinebecker, Hölzel
Arbuscular mycorrhizal fungal root colonization		Cultured in sterile soil in the field from April to October 2011 and then extracted with sodium hexametaphosphate (35 g l^{-1}). Hyphal length was quantified after staining with trypan blue.	Morris, Rillig
Stability of soil aggregates		A subsample of the same soil than above (AMF colonization) was passed through a 250 µm sieve under water to determine the percentage of water stable macroaggregates.	Morris, Rillig
Soil organic Carbon		Measured in 14 soil cores (0-10 cm). Calculated as the difference between total carbon (measured with a CN analyzer "Vario Max" [Elementar Analysensysteme GmbH, Hanau, Germany]) and inorganic carbon (determined after combustion of organic carbon in a muffle furnace; 450°C for 16 h), done in samples collected in May 2011	Schöning, Solly, Schrumpf
Forage quality		Was calculated as a function of mean of scaled crude protein concentration and scaled relative forage value, done in May-June in 2008-2012.	Klaus, Kleinebecker, Hölzel
Resistance to plant pathogens		Calculated as the inverse of the total cover of foliar fungal pathogens. The cover of pathogens was measured in four 25 × 1 m transects per plot, were proportion of plants infected, and leaf area infected of these individuals was measured; done in October 2011,	Blaser, Prati, Fischer
Pest control		Number of trap nesting wasps known to feed on pest insects, done between April and October 2008.	Steckel, Westphal, Steffan-Dewenter
Pollinator abundance		Estimated as the total abundance of flower visitors, measured in one 200 × 3 m transect per plot, done in May 2008	Krauss, Klein, Weiner, Werner, Blüthgen
Bird diversity		Measured as the cumulative species richness estimated by audio-visual point-counts, done in May-June 2008-2010	Renner, Böhm, Tschapka
Flower cover		Measured as the number of inflorescences in four 50 × 3 m transects per plot. Flower area for each species was obtained from the literature.	Binkenstein, Schaefer

Note that for some groups the taxonomic unit was either operational taxonomic units (OTU: fungi and protists) or families (bacteria and below-ground insect larvae). Abundance measures were: per cent cover (plants, bryophytes), number of individuals captured (arthropods) and relative proportion of sequence reads assigned to each family among all reads within each plot (protists, soil bacteria and mycorrhiza).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 3 | Correlations between diversity predictors from the models in the main text

Models using residuals after controlling for environmental predictors

		Belowground herbivores	Belowground predators	Detritivores	Aboveground herbivores	Soil microbial decomposers	Bacterivores	Aboveground predators	Plant symbionts	Primary producers	Belowground herbivores	Belowground predators	Detritivores	Aboveground herbivores	Soil microbial decomposers	Bacterivores	Aboveground predators	Plant symbionts
Abundance	Primary producers	-0.06	0.09	0.03	-0.07	-0.01	0.01	-0.13	-0.05	0.06	0.06	-0.05	0.05	-0.06	0.10	-0.02	-0.04	-0.16
	Belowground herbivores		0.21	0.08	0.19	0.00	0.18	-0.02	-0.02	-0.15	0.37	0.10	0.13	0.08	0.06	0.18	-0.05	0.10
	Belowground predators			0.14	-0.16	0.05	0.09	-0.23	-0.04	-0.16	0.22	0.54	0.12	0.09	0.20	0.15	-0.09	0.12
	Detritivores				0.12	-0.10	-0.02	0.20	-0.02	0.04	0.15	0.09	0.74	0.12	-0.13	0.02	0.03	-0.15
	Aboveground herbivores					0.01	0.16	0.32	-0.05	-0.05	0.05	0.08	0.22	0.54	-0.04	0.14	0.34	-0.06
	Soil microbial decomposers						0.27	-0.02	0.05	0.00	0.01	0.04	-0.05	-0.11	0.17	0.25	-0.05	0.08
	Bacterivores							-0.13	0.06	-0.24	0.11	0.12	0.00	-0.10	0.13	0.76	-0.14	0.25
	Aboveground predators								-0.04	0.14	0.02	-0.11	0.14	0.38	-0.16	-0.18	0.59	-0.18
	Plant symbionts									-0.06	0.03	-0.10	0.04	-0.12	0.03	-0.06	0.03	0.46
Richness	Primary producers										-0.13	-0.13	-0.09	0.16	-0.38	-0.26	0.02	-0.22
	Belowground herbivores											0.24	0.18	0.11	-0.03	0.09	0.11	0.13
	Belowground predators												0.08	0.13	0.19	0.25	-0.02	-0.07
	Detritivores													0.19	-0.12	0.02	0.09	-0.08
	Aboveground herbivores														0.02	-0.03	0.48	-0.15
	Soil microbial decomposers															0.21	-0.06	0.17
	Bacterivores																-0.17	0.21
	Aboveground predators																	-0.17

Models including environmental predictors

		Soil depth	Fertilisation	Mowing	Grazing	Elevation	TWI	Primary producers	Belowground herbivores	Belowground predators	Detritivores	Aboveground herbivores	Soil microbial decomposers	Bacterivores	Aboveground predators	Plant symbionts	Primary producers	Belowground herbivores	Belowground predators	Detritivores	Aboveground herbivores	Soil microbial decomposers	Bacterivores	Aboveground predators	Plant symbionts
Environment	pH	0.05	0.07	0.05	-0.17	-0.24	0.22	-0.19	-0.22	-0.09	0.12	-0.21	-0.40	-0.30	0.25	-0.03	0.08	-0.18	0.00	0.21	-0.07	-0.05	0.21	-0.07	
	Soil depth		-0.21	0.02	-0.01	-0.82	0.52	-0.28	-0.22	-0.11	0.22	-0.29	-0.15	-0.06	-0.10	0.60	-0.62	-0.16	0.20	0.26	-0.39	0.42	0.10	-0.02	0.37
	Fertilisation			0.61	-0.22	0.15	0.00	0.01	-0.06	0.22	0.06	-0.06	0.00	0.10	0.01	-0.16	-0.23	-0.04	0.09	0.13	-0.06	-0.08	-0.01	-0.10	-0.06
	Mowing				-0.68	0.05	0.18	-0.03	-0.09	0.19	0.09	-0.15	-0.17	-0.01	-0.04	-0.12	-0.40	-0.13	0.10	0.09	-0.24	-0.02	-0.04	-0.17	-0.16
	Grazing					-0.09	-0.13	0.01	0.04	-0.02	-0.04	0.16	0.26	0.27	0.01	0.11	0.13	0.12	0.14	-0.04	0.11	0.07	0.32	0.12	0.25
	Elevation						-0.55	0.31	0.18	0.11	-0.28	0.29	0.22	0.05	-0.04	-0.61	0.60	0.19	-0.20	-0.33	0.39	-0.32	-0.13	-0.10	-0.42
	TWI							-0.19	-0.20	-0.02	0.15	-0.28	-0.36	-0.19	0.09	0.35	-0.50	-0.11	0.19	0.27	-0.28	0.41	-0.08	0.14	0.10
Abundance	Primary producers								-0.06	0.13	-0.06	0.22	0.06	-0.08	-0.27	-0.21	0.12	0.06	-0.15	-0.11	0.10	-0.10	-0.13	-0.14	-0.41
	Belowground herbivores									0.26	-0.02	0.11	0.26	0.28	-0.05	-0.08	0.19	0.48	0.08	-0.03	0.22	-0.09	0.24	-0.12	0.20
	Belowground predators										0.09	-0.01	0.08	0.14	-0.19	-0.09	-0.08	0.24	0.51	0.06	0.09	0.09	0.13	-0.13	0.05
	Detritivores											0.02	-0.15	-0.05	0.16	0.15	-0.16	0.09	0.20	0.88	0.01	-0.06	0.02	0.08	0.03
	Aboveground herbivores												0.07	0.06	0.23	-0.25	0.17	0.05	-0.06	-0.01	0.64	-0.14	0.03	0.31	-0.29
	Soil microbial decomposers													0.47	-0.18	0.05	0.17	0.17	0.01	-0.21	0.09	-0.11	0.46	-0.18	0.32
	Bacterivores														-0.17	0.05	-0.07	0.15	0.12	-0.08	-0.07	0.08	0.82	-0.20	0.40
	Aboveground predators															-0.11	0.23	-0.03	-0.09	0.18	0.32	-0.13	-0.15	0.68	-0.07
	Plant symbionts																-0.38	0.01	0.06	0.19	-0.29	0.28	0.13	-0.05	0.64
Richness	Primary producers																	0.14	-0.22	-0.24	0.42	-0.31	-0.13	0.14	-0.14
	Belowground herbivores																		0.22	0.08	0.18	-0.08	0.11	0.03	0.19
	Belowground predators																			0.17	-0.01	0.27	0.20	-0.02	0.12
	Detritivores																				0.04	-0.01	-0.05	0.15	0.08
	Aboveground herbivores																					-0.17	-0.09	0.41	-0.22
	Soil microbial decomposers																						0.14	-0.04	0.24
	Bacterivores																							-0.18	0.46
	Aboveground predators																								-0.09

Correlations between residuals (after controlling for the effect of study region, soil type, pH, topographic wetness index and the three land-use intensity components: fertilization, mowing and grazing) of abundance and species richness of the nine different trophic groups considered (top) or of the raw data (bottom). Those predictors removed owing to multicollinearity problems are shaded grey, with the correlation responsible highlighted. TWI = topographic wetness index, obtained from P. M., unpublished data.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 4 | Model selection

		Biomass	Forage quality	Potential nitrification	Root biomass	Root decomposition	Phophorus retention	Mycorrhizal colonization	Soil.aggregate stability	Soil C	Pest control	Resistance to pathogen	Pollinator abundance	Bird diversity	Flower cover	Multifunctionality 25%	Multifunctionality 50%	Multifunctionality 25%	Multifunctionality 90%
Abundance	Primary producers	3.26	-1.92	-1.09	-1.08	-2.34	0.59	-1.90	-2.00	-1.48	-0.35	6.09	-1.84	-1.99	3.53	-1.48	-1.76	-1.48	-1.69
	Belowground herbivores	-1.85	-0.50	1.35	-2.00	-1.88	-0.13	-0.02	-1.99	-1.70	-2.00	1.88	0.13	-0.10	-1.24	-0.65	-2.00	-0.98	-1.95
	Belowground predators	-1.90	0.03	-1.95	5.88	-1.97	-1.75	3.10	-0.70	-2.00	-1.74	-1.62	-1.71	-0.19	-1.98	-1.90	-1.34	-0.30	-0.06
	Aboveground herbivores	-0.74	-1.71	11.71	0.94	-0.46	-1.96	4.83	3.71	3.26	-1.96	-0.77	-1.97	-2.00	1.29	1.46	-0.03	-0.63	-1.74
	Soil microbial decomposer	-1.57	-0.86	0.38	-0.20	-0.14	-1.29	5.10	7.13	1.45	-1.99	0.66	-0.75	1.55	-1.99	-2.00	1.68	-1.98	-1.80
	Aboveground predators	0.74	-1.25	1.18	4.38	-1.84	-1.88	-1.94	2.68	0.48	1.79	6.41	-1.97	2.10	-1.81	-0.25	0.59	2.23	-1.28
	Plant symbionts	-0.80	-1.22	5.99	0.55	-0.39	1.82	-0.93	1.39	0.03	0.49	-1.96	2.92	-1.85	-1.96	-1.87	-0.97	-1.83	-2.00
Richness	Primary producers	14.90	5.43	12.00	1.94	0.13	16.91	-0.32	4.28	-0.97	5.01	1.56	-1.74	33.27	-1.66	-1.83	-1.35	6.88	2.45
	Belowground herbivores	-0.09	-1.95	-1.93	-0.23	-1.92	-1.94	-1.99	2.13	-1.61	-0.71	1.13	-2.00	-1.22	-1.99	-1.51	-1.88	-0.61	-1.74
	Belowground predators	-0.08	-1.78	-0.59	-1.65	-1.97	-0.56	-1.83	-1.65	-0.25	-1.59	-0.04	-1.84	-1.98	-1.02	-1.76	-1.53	-1.47	0.04
	Detritivores	0.74	-1.89	2.56	-1.79	-1.63	-1.22	-1.15	-1.35	-2.00	-1.32	-1.68	-1.63	-1.63	-0.44	-1.32	-1.66	-1.70	-1.64
	Aboveground herbivores	5.68	5.14	-1.85	6.91	-1.99	-1.67	-1.66	-1.66	0.10	-1.94	-0.65	-1.57	5.90	6.40	-1.49	-0.76	-1.98	-1.98
	Soil microbial decomposer	-1.97	-0.80	0.47	-3.86	-1.77	-1.20	4.95	-1.41	-1.37	-1.99	-1.54	-1.66	-1.66	-1.63	-1.92	0.01	-0.43	-0.77
	Bacterivores	-1.93	2.98	-0.86	-1.55	-1.80	0.74	-1.84	3.58	-1.75	-1.58	-1.91	1.44	-1.87	-1.90	-1.56	-1.33	-1.05	-1.90
	Aboveground predators	2.31	-1.93	-1.43	5.79	-1.81	10.25	-0.23	-1.97	6.17	3.29	-0.75	-1.38	-1.88	-1.34	-0.53	-2.00	-0.48	-1.60
	Plant symbionts	-1.91	-1.87	-1.91	-0.35	-1.58	-1.36	-1.39	-1.38	0.55	-1.99	2.64	7.34	-2.00	-2.00	-1.32	-2.00	-0.03	1.27

Difference in AIC when subtracting each term regarding the full model according to the backward step AIC procedure used (models using the environmental-corrected residuals, as presented in Fig. 1 and Extended Data Fig. 1). Green shade indicates the terms included in the most parsimonious models. Orange shade indicates terms included in the model with the lowest AIC but further removed using *F*-ratio tests.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

LETTER

doi:10.1038/nature19074

The TRPM2 ion channel is required for sensitivity to warmth

Chun-Hsiang Tan[1,2] & Peter A. McNaughton[1]

Thermally activated ion channels are known to detect the entire thermal range from extreme heat (TRPV2), painful heat (TRPV1, TRPM3 and ANO1), non-painful warmth (TRPV3 and TRPV4) and non-painful coolness (TRPM8) through to painful cold (TRPA1)[1–7]. Genetic deletion of each of these ion channels, however, has only modest effects on thermal behaviour in mice[6–12], with the exception of TRPM8, the deletion of which has marked effects on the perception of moderate coolness in the range 10–25 °C[13]. The molecular mechanism responsible for detecting non-painful warmth, in particular, is unresolved. Here we used calcium imaging to identify a population of thermally sensitive somatosensory neurons which do not express any of the known thermally activated TRP channels. We then used a combination of calcium imaging, electrophysiology and RNA sequencing to show that the ion channel generating heat sensitivity in these neurons is TRPM2. Autonomic neurons, usually thought of as exclusively motor, also express TRPM2 and respond directly to heat. Mice in which TRPM2 had been genetically deleted showed a striking deficit in their sensation of non-noxious warm temperatures, consistent with the idea that TRPM2 initiates a 'warm' signal which drives cool-seeking behaviour.

Previous studies have described novel heat-sensitive neurons not activated by agonists for any of the known heat-sensitive ion channels[5,14–16]. We identified these neurons in cultures from dorsal root ganglia (DRG) by using calcium imaging and selective agonists for known thermo-TRP ion channels. We found that around 10% of neurons responded to heat, but were not activated by any of the agonists for known TRP channels (Fig. 1a–c and Supplementary Video 1). Changing the order of application of agonists or using a lower starting temperature had little effect on this proportion (Extended Data Figs 1 and 2). Novel heat-sensitive neurons were found to be activated over a wide range of temperatures, with a subset activated in the range of warm temperatures between 34 °C and 42 °C, suggesting a possible role in warmth sensation, and a second group activated only at higher temperatures (Fig. 1d, e). Novel heat-sensitive neurons were significantly larger than either TRPV1-expressing or TRPM3-expressing neurons (Extended Data Fig. 3). We investigated the possibility that the novel heat-sensitivity may be co-expressed with TRPV1 and TRPM3 by blocking these channels with antagonists before applying heat; the higher proportion of neurons responding (46% in Extended Data Fig. 4a, compared to <10% expressing only the novel heat-sensitive mechanism, see Fig. 1 and Extended Data Figs 1, 2 and 4b) shows that there is significant co-expression of the novel heat-sensitive mechanism with TRPV1 and TRPM3. The phenotype of neurons expressing the novel heat-sensitive mechanism was investigated by identifying novel heat-sensitive neurons in the presence of TRP channel blockers (Extended Data Fig. 5a, red) and then exposing them to IB4, which marks a non-peptidergic neuronal population (Extended Data Fig. 5b, green). These results show that the novel heat-sensitive mechanism is predominantly expressed in $IB4^+$ neurons (74% $IB4^+$).

We next sought to identify a source of neurons which expresses a less-complex set of heat-sensitive ion channels than are present in DRG neurons. In isolated sympathetic neurons from the superior cervical ganglion (SCG) we found that no neuron showed an increase in intracellular calcium concentration ($[Ca^{2+}]_i$) in response to agonists of known thermo-TRP channels, but that 58% showed a significant response to heat (Fig. 2a and Extended Data Fig. 6). Similar results were obtained in parasympathetic neurons isolated from the pterygopalatine ganglion (PPG), in which 49% of neurons showed novel heat sensitivity (not shown). Autonomic neurons therefore express the novel heat-activated ion channel in isolation, which offers an advantageous preparation for determining its properties.

We found that the heat-activated calcium increase in autonomic neurons was due to an influx of Ca^{2+} from the external solution, and that it was reduced but not abolished by removal of external Na^+ (Extended Data Fig. 6c). The TRPV channel family activator 2-APB[17] suppressed the heat-activated Ca^{2+} influx, and the TRPV blocker ruthenium red[2] had no effect (Extended Data Fig. 6d, e), making it unlikely that the novel heat-activated mechanism is a TRPV family member. The Ca^{2+} influx was unaffected by the voltage-dependent Na channel blocker tetrodotoxin (TTX) at a high enough concentration to block the TTX-insensitive Na^+ channels $Na_V1.8$ and $Na_V1.9$ (Extended Data Fig. 6f). L-type Ca^{2+} channel blockers prevented firing of action potentials during simultaneous recordings of membrane voltage and intracellular calcium imaging, but did not completely block the heat-activated Ca^{2+} influx (Fig. 2b; see also Extended Data Fig. 6g). Together these observations show that the heat-sensitive ion channel is permeable both to Ca^{2+} and to Na^+, and that when the channel is opened by heat, the resulting depolarization activates L-type calcium channels and thus augments the calcium influx. The fact that membrane current through the novel heat-activated ion channel is carried by Na^+ and Ca^{2+} makes it unlikely that the channel is ANO1, which is permeable to chloride ions[6].

We next investigated the voltage- and time-dependent behaviour of the novel heat-activated ion channel, isolated by suppressing voltage-dependent calcium, sodium and potassium currents[1]. The current-voltage relation was approximately linear, with a reversal potential close to zero (Fig. 2c), and the channel showed no time-dependent gating by membrane voltage (Extended Data Fig. 7). Interestingly, activation of the channel by mild warmth was strongly potentiated by hydrogen peroxide, both in autonomic neurons (Fig. 2d) and in DRG neurons (Extended Data Fig. 4c, d).

To identify the ion channel responsible for the novel heat sensitivity we carried out an analysis of mRNA expression by the use of RNA sequencing (RNA-seq). We investigated two cell lines, MAH cells[18] and PC12 cells, both derived from rat adrenal cells, which share many properties with sympathetic neurons. Like primary autonomic neurons, no cell in either line responded to agonists of any of the conventional thermo-TRP channels, but a significant fraction of cells responded to heat with thermal thresholds similar to those in somatosensory neurons

[1]Wolfson Centre for Age-Related Diseases, King's College London, Guy's Campus, London Bridge, London SE1 1UL, UK. [2]Department of Neurology, Kaohsiung Medical University Hospital, Kaohsiung Medical University, Kaohsiung, Taiwan.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

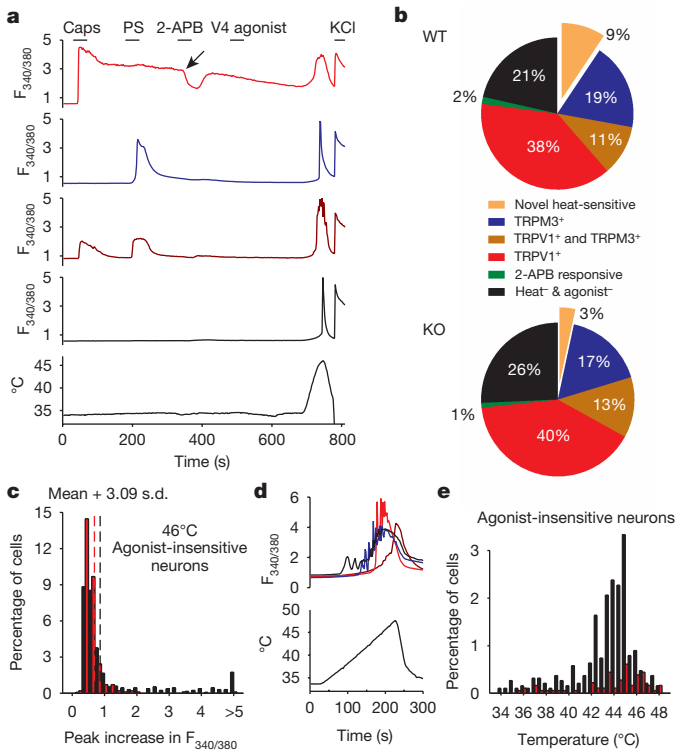


Figure 1 | Around 10% of somatosensory neurons demonstrate a novel heat-sensitive response. a, Examples of increases of $[Ca^{2+}]_i$ measured with fura-2 ($F_{340/380}$ ratio, ordinate) in response to known TRP channel agonists and to heat. From top: TRPV1-expressing neuron responding to capsaicin (caps, red); TRPM3-expressing neuron responding to pregenolone sulfate (PS, blue); TRPV1-and TRPM3-co-expressing neuron (brown); neuron unresponsive to TRP channel agonists but showing a response to heat (bottom trace). See Supplementary Information Video 1. **b**, Percentages of neurons expressing TRPV1, TRPM3 and the novel heat-sensitive mechanism in DRG neurons from wild-type (WT) (top) and *Trpm2*$^{-/-}$ knockout (KO) mice (bottom). Green: neurons responding to 2-APB but not to other agonists. **c**, Histogram of maximal $F_{340/380}$ ratio increase in neurons insensitive to TRP agonists. Black, wild-type neurons; red, *Trpm2*$^{-/-}$ neurons. Vertical dashed lines: thresholds discriminating between heat-sensitive and insensitive neurons (see Extended Data Fig. 1). **d**, Thermal thresholds of novel heat-sensitive DRG neurons. Increases of $F_{340/380}$ ratio (top) in response to slowly rising heat ramp (bottom). **e**, Temperature thresholds of novel heat-sensitive neurons from wild-type (black) and *Trpm2*$^{-/-}$ mice (red). Proportion in range 34–42 °C reduced from 4.4% in wild type to 0.9% in *Trpm2*$^{-/-}$ ($P \leq 0.0001$; Fisher's exact test), and in range 42–48 °C from 15% in wild type (290/1,890) to 3% in *Trpm2*$^{-/-}$ (55/1,800) ($P \leq 0.0001$; Fisher's exact test). Cell numbers and replicates for **a**–**c** were 1,324 DRG neurons from one wild-type mouse on 4 coverslips and 981 DRG neurons from one *Trpm2*$^{-/-}$ mouse on 4 coverslips were imaged. Similar results obtained using 52 additional coverslips from 9 additional wild-type mice and 10 coverslips from 3 additional *Trpm2*$^{-/-}$ mice; some of these results are shown in Extended Data Figs 1, 2 and 4. Cell numbers and replicates for **d** and **e** were 1,890 DRG neurons from one wild-type mouse on 8 coverslips and 1,800 DRG neurons from one *Trpm2*$^{-/-}$ mouse on 8 coverslips were imaged. Similar results obtained using 10 coverslips from one additional wild-type mouse.

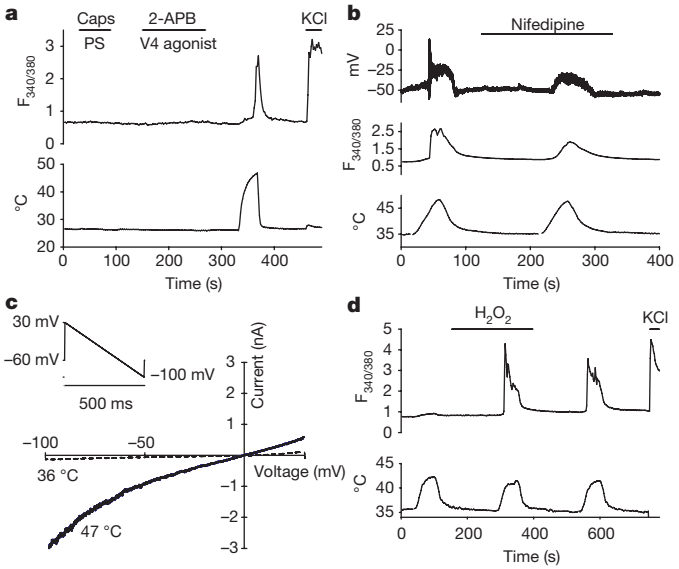


Figure 2 | Properties of the novel heat-sensitive ion channel in autonomic neurons. a, Sympathetic neurons from superior cervical ganglion (SCG) respond to heat but not to TRP channel agonists. **b**, The L-type Ca^{2+} channel blocker nifedipine (10 μM) blocks spiking but not steady depolarization in response to heat in a patch-clamped PPG neuron (top), and reduces, but does not eliminate the Ca^{2+} increase (middle). Simultaneous recording of membrane potential and $[Ca^{2+}]_i$ ($F_{340/380}$ ratio) in current-clamped PPG neuron. **c**, Current–voltage relations at 36 °C and 47 °C of voltage-clamped PPG neuron in response to voltage ramp shown in the top left. Similar I/V relation observed with inverse voltage ramp starting from −100 mV (Extended Data Fig. 7). See Methods for details. **d**, H_2O_2 (400 μM) potentiates Ca^{2+} increase in response to a mild temperature stimulus (42 °C) in SCG neurons. Note potentiation is long-lasting after H_2O_2 is removed. Similar results obtained in PPG neurons. Percentage of neurons ($n = 456$) responding to heat, determined as in Extended Data Fig. 1b, was 5% before, 59% during and 54% post-H_2O_2, respectively. Cell numbers and replicates for **a** were 166 SCG neurons from 3 wild-type mice on 3 coverslips imaged for this experiment. Similar results obtained with SCG neurons using 15 additional coverslips from 9 additional mice and with PPG neurons using 14 coverslips from 9 additional mice. Cell numbers and replicates for **b** were 2 PPG neurons on 2 coverslips simultaneously patch-clamped and imaged and these showed a similar response. Cell numbers and replicates for **c** were 3 PPG neurons on 3 coverslips simultaneously patch-clamped and imaged and these showed a similar response. Cell numbers and replicates for **d** were 456 SCG neurons from two wild-type mice on 4 coverslips imaged and these showed a similar enhancement with H_2O_2; 731 PPG neurons (not shown) from three wild-type mice on 3 coverslips were imaged and showed a similar enhancement to the SCG neurons. Similar results obtained with SCG neurons using 9 coverslips from 5 additional mice and with PPG neurons using 3 coverslips from 3 additional mice.

(Extended Data Fig. 8). We also found that the fraction of heat-sensitive neurons in both lines was reduced by differentiation to a neuronal-like phenotype (Extended Data Fig. 8, red bars).

The properties of mixed Na^+/Ca^{2+} permeability, a reversal potential near 0 mV and absence of time-dependent gating by membrane potential (see earlier) are consistent with a member of the large TRP and CNG ion channel families. RNA-seq analysis of the MAH cell line showed that detectable mRNA was present only for the seven TRP channels shown in Table 1, out of all TRP and CNG channels. TRPC3, TRPV2, TRPM4 and TRPM7 are unlikely candidates for the warmth-sensitive channel as all have nonlinear current–voltage (I/V) relations[19], whereas the warmth-sensitive ion channel is linear (Fig. 2c). TRPC1 and TRPV2 are strongly upregulated by culture in differentiation medium (Table 1), which contrasts with the downregulation observed for the heat-sensitive ion channel (Extended Data Fig. 8a). TRPV2 is activated only by extreme heat[20] and TRPM4 is calcium-impermeable[21], neither of which is consistent with the properties of the novel heat-sensitive ion channel (see above). *TRPC2* is a pseudogene in primates and seems unlikely to play an important role in behavioural warmth sensation, a fundamental property in all mammals. This leaves only TRPM2, which has a linear I/V relation (see Fig. 2c)[22], is activated at temperatures above 35 °C[22] and is expressed in DRG neurons[23,24], as the most probable candidate for the warmth-sensitive ion channel. Thermal activation of TRPM2 is enhanced by hydrogen peroxide[25] and the channel is blocked by the chemical 2-APB[26], both of which are characteristics shared by the novel heat-sensitive ion channel (Fig. 2d and Extended Data Figs 4c, d and 6d).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Table 1 | mRNA levels of TRP ion channel genes in MAH cells

Gene	FPKM in growth medium	FPKM in differentiation medium	Log_2 fold change	*P* value	*q* value
Trpc1	1.779	4.700	1.401	<0.001	<0.001
Trpc2	6.804	4.995	−0.446	0.121	0.323
Trpc3	5.641	1.309	−5.729	<0.001	<0.001
Trpv2	0.035	4.003	6.840	<0.001	<0.001
Trpm2	1.017	0.970	−0.069	0.800	0.922
Trpm4	0.277	0.400	0.532	0.112	0.310
Trpm7	18.108	15.428	−0.231	0.196	0.428

All genes within the TRP channel family and the six cyclic nucleotide-regulated channels (*Cnga1, Cnga2, Cnga3, Cnga4, Cngb1, and Cngb3*) were examined. The seven genes listed are the ones whose mRNA levels are detectable in MAH cells in either condition. FPKM, fragments per kilobase of transcript per million mapped reads. The *P* values represent significance of change of expression using one-way ANOVA; *q* values are post-hoc corrected for multiple testing using the false discovery rate (FDR) method. Two biological replicates in each condition were deep-sequenced. Growth medium contained dexamethasone (5 μM), differentiation medium contained the growth factors bFGF (10 ng ml^{-1}), CNTF (10 ng ml^{-1}) and NGF (50 ng ml^{-1}).

We confirmed the identity of the novel heat-sensitive ion channel using calcium imaging experiments on neurons from *Trpm2*$^{-/-}$ mice[27,28]. The proportions of DRG neurons responding to TRP channel agonists were similar to those in wild-type neurons (Fig. 1b and Extended Data Figs 1c and 2c), but the proportion of novel heat-sensitive neurons was significantly reduced, from ~10% to around 3% (Fig. 1c, red bars; $P \leq 0.0001$, Fisher's exact test, see also Extended Data Figs 1, 2 and 4). Moreover, the mean amplitude of the increase in the fura-2 fluorescence ratio at 340 and 380 nm (an index of the increase in intracellular calcium) in *Trpm2*$^{-/-}$ neurons not responding to TRP channel agonists was greatly reduced, from $F_{340/380} = 2.336 \pm 0.2060$ in wild type, to 0.6202 ± 0.2435 in *Trpm2*$^{-/-}$ (Fig. 1c; $P \leq 0.0001$, two-tailed unpaired *t*-test). Notably, the sensitivity of the heat response to hydrogen peroxide was abolished by deletion of *Trpm2* (Extended Data Fig. 4c, d). In experiments on the thermal thresholds of novel heat-sensitive neurons (Fig. 1d, e), deletion of *Trpm2* abolished almost all thermal sensitivity in the range 34–42 °C (reduced from 4.4% in wild type to 0.9% in *Trpm2*$^{-/-}$, $P \leq 0.0001$; Fisher's exact test), though some neurons still responded to temperatures in the noxious thermal range (reduced from 15% in wild type to 3.1% in *Trpm2*$^{-/-}$, $P \leq 0.0001$; Fisher's exact test). We also considered the possibility that TRPM2 may contribute to the heat response in DRG neurons in which it is co-expressed with TRPV1 and/or TRPM3. Deletion of *Trpm2* was found not to affect the maximum amplitude of the response to heat in TRPV1-expressing neurons, but it did reduce the response of TRPM3-expressing neurons (Extended Data Fig. 9), suggesting that co-activation of TRPM2 and TRPM3 is important in determining the heat responses of these neurons. Finally, in SCG neurons from *Trpm2*$^{-/-}$ mice, both the numbers and response amplitudes of heat-sensitive neurons were greatly reduced (Extended Data Fig. 6b).

Expression of mRNA that encodes TRPM2 was demonstrated in heat-sensitive DRG neurons using *in situ* hybridization. *Trpm2* mRNA was surprisingly abundant in neurons, being expressed in 89% of DRG neurons, but was expressed in, at most, a very small minority of glial cells (Extended Data Fig. 10). In prior tests for heat sensitivity in the same neurons, 42% of neurons positive for *Trpm2* mRNA had responded when TRPV1 and TRPM3 were blocked. The reason why not all neurons positive for *Trpm2* mRNA respond to heat is not clear, but could be due to a low conversion of mRNA into TRPM2 protein or to low trafficking to the membrane in around half of DRG neurons. However, many fewer neurons responded to heat in the population not expressing mRNA for *Trpm2* (13%, see Extended Data Fig. 10), which is consistent with the small number of heat-sensitive neurons still seen in DRG cultures from Trpm2$^{-/-}$ mice (see for example, Fig. 1b).

Finally, we investigated whether genetic deletion of *Trpm2* has an impact on thermal preference in mice. The most striking behavioural difference was that wild-type mice avoided the non-noxious warm temperature of 38 °C, whereas *Trpm2*$^{-/-}$ mice showed little preference (Fig. 3c, e and Supplementary Video 2). The difference became much less noticeable at 43 °C, when noxious-heat avoidance mechanisms

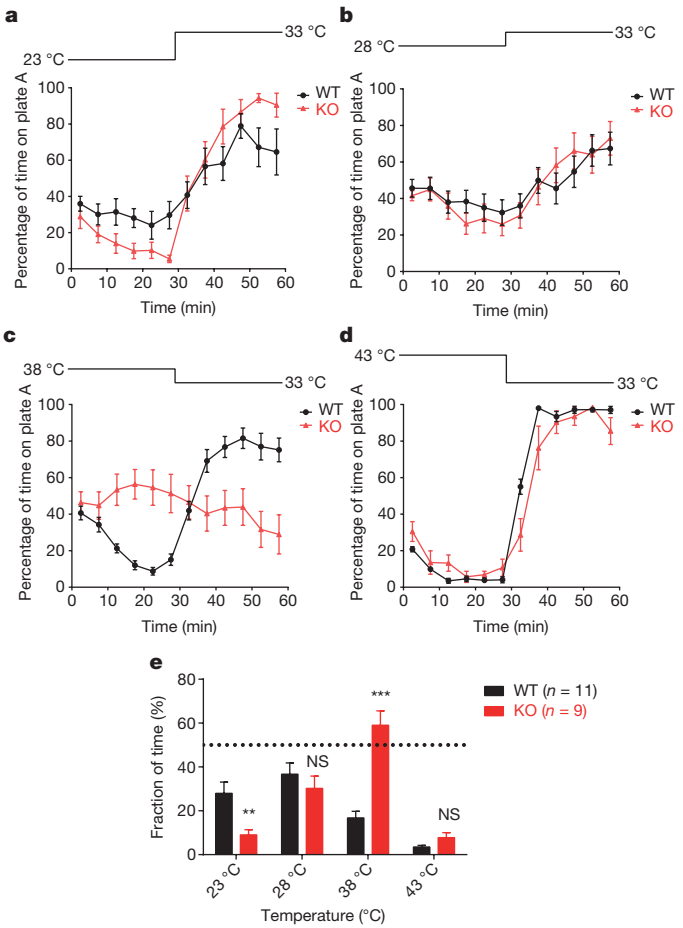


Figure 3 | Deletion of TRPM2 shifts adult male mouse thermal preference towards warmer temperatures. a–d, Two-plate thermal preference test. One plate is at 33 °C at the start of the experiment, the other ('plate A') is at a variable temperature as shown above each graph. Temperature reversed at $t = 30$ min to account for any possible bias caused by external cues. Points show mean behavioural preference averaged over 5 min intervals (error bars represent the mean ± s.e.m., $n = 11$ for wild type and 9 for *Trpm2*$^{-/-}$). Mice were wild-type or *Trpm2*$^{-/-}$ (KO) male littermates, aged 12–16 weeks of age. No difference was observed between the thermal behaviour of wild-type and heterozygous *Trpm2*$^{+/-}$ mice (not shown, $n = 4$). **e**, Mean thermal preference averaged from the data between $t = 15$–30 min and $t = 45$–60 min shown in **a–d**. Error bars represent the mean ± s.e.m. $^{**}P < 0.01$; $^{***}P < 0.001$; NS, $P > 0.05$; unpaired *t*-test. Similar results were obtained with experiments using non-littermates in which 12 wild-type male mice from Charles River were compared with 7 *Trpm2*$^{-/-}$ male mice from homozygote breeding pairs.

driven by TRPV1, TRPM3 and ANO1 become important[3–9]. Wild-type mice also showed a less strong aversion than *Trpm2*$^{-/-}$ mice for the non-noxious cool temperature of 23 °C (Fig. 3a, e). Expression of TRPM2 therefore causes wild-type mice to prefer cooler temperatures over a range of temperatures extending from 23 °C to above 38 °C, though we note that the actual temperature at the sensory nerve ending will be higher than the plate temperature, particularly at the lower end of this range, because of the influence of body temperature.

The work presented here shows that TRPM2 accounts for a novel heat-sensitive mechanism in both somatosensory and autonomic neurons. The altered thermal preference in *Trpm2*$^{-/-}$ mice supports the hypothesis that TRPM2 expressed in somatosensory neurons provides a non-noxious 'warm' signal which drives mice to seek cooler temperatures over a wide temperature range, from 23 °C to above 38 °C. Other studies have shown that TRPM8 provides a 'cool' signal which drives warmth-seeking behaviour over the range 10–25 °C[13], whereas activation of TRPV1, TRPM3 and ANO1 provide a high-temperature noxious-heat-avoidance signal[3–9].

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 June 2015; accepted 1 July 2016.
Published online 17 August 2016.

1. Cesare, P. & McNaughton, P. A novel heat-activated current in nociceptive neurons and its sensitization by bradykinin. *Proc. Natl Acad. Sci. USA* **93,** 15435–15439 (1996).
2. Caterina, M. J. *et al.* The capsaicin receptor: a heat-activated ion channel in the pain pathway. *Nature* **389,** 816–824 (1997).
3. Jordt, S. E., McKemy, D. D. & Julius, D. Lessons from peppers and peppermint: the molecular logic of thermosensation. *Curr. Opin. Neurobiol.* **13,** 487–492 (2003).
4. Basbaum, A. I., Bautista, D. M., Scherrer, G. & Julius, D. Cellular and molecular mechanisms of pain. *Cell* **139,** 267–284 (2009).
5. Vriens, J. *et al.* TRPM3 is a nociceptor channel involved in the detection of noxious heat. *Neuron* **70,** 482–494 (2011).
6. Cho, H. *et al.* The calcium-activated chloride channel anoctamin 1 acts as a heat sensor in nociceptive neurons. *Nature Neurosci.* **15,** 1015–1021 (2012).
7. Vriens, J., Nilius, B. & Voets, T. Peripheral thermosensation in mammals. *Nature Rev. Neurosci.* **15,** 573–589 (2014).
8. Caterina, M. J. *et al.* Impaired nociception and pain sensation in mice lacking the capsaicin receptor. *Science* **288,** 306–313 (2000).
9. Davis, J. B. *et al.* Vanilloid receptor-1 is essential for inflammatory thermal hyperalgesia. *Nature* **405,** 183–187 (2000).
10. Moqrich, A. *et al.* Impaired thermosensation in mice lacking TRPV3, a heat and camphor sensor in the skin. *Science* **307,** 1468–1472 (2005).
11. Huang, S. M., Li, X., Yu, Y., Wang, J. & Caterina, M. J. TRPV3 and TRPV4 ion channels are not major contributors to mouse heat sensation. *Mol. Pain* **7,** 37 (2011).
12. Miyamoto, T., Petrus, M. J., Dubin, A. E. & Patapoutian, A. TRPV3 regulates nitric oxide synthase-independent nitric oxide synthesis in the skin. *Nature Commun.* **2,** 369 (2011).
13. Bautista, D. M. *et al.* The menthol receptor TRPM8 is the principal detector of environmental cold. *Nature* **448,** 204–208 (2007).
14. Nagy, I. & Rang, H. Noxious heat activates all capsaicin-sensitive and also a sub-population of capsaicin-insensitive dorsal root ganglion neurons. *Neuroscience* **88,** 995–997 (1999).
15. Woodbury, C. J. *et al.* Nociceptors lacking TRPV1 and TRPV2 have normal heat responses. *J. Neurosci.* **24,** 6410–6415 (2004).
16. Lawson, J. J., McIlwrath, S. L., Woodbury, C. J., Davis, B. M. & Koerber, H. R. TRPV1 unlike TRPV2 is restricted to a subset of mechanically insensitive cutaneous nociceptors responding to heat. *J. Pain* **9,** 298–308 (2008).
17. Hu, H. Z. *et al.* 2-aminoethoxydiphenyl borate is a common activator of TRPV1, TRPV2, and TRPV3. *J. Biol. Chem.* **279,** 35741–35748 (2004).
18. Birren, S. J. & Anderson, D. J. A v-*myc*-immortalized sympathoadrenal progenitor cell line in which neuronal differentiation is initiated by FGF but not NGF. *Neuron* **4,** 189–201 (1990).
19. Wu, L.-J., Sweet, T.-B. & Clapham, D. E. International Union of Basic and Clinical Pharmacology. LXXVI. Current progress in the mammalian TRP ion channel family. *Pharmacol. Rev.* **62,** 381–404 (2010).
20. Caterina, M. J., Rosen, T. A., Tominaga, M., Brake, A. J. & Julius, D. A capsaicin-receptor homologue with a high threshold for noxious heat. *Nature* **398,** 436–441 (1999).
21. Vennekens, R. & Nilius, B. Insights into TRPM4 function, regulation and physiological role. *Handb. Exp. Pharmacol.* **179,** 269–285 (2007).
22. Togashi, K. *et al.* TRPM2 activation by cyclic ADP-ribose at body temperature is involved in insulin secretion. *EMBO J.* **25,** 1804–1815 (2006).
23. Naziroglu, M., Ozgul, C., Celik, O., Cig, B. & Sozbir, E. Aminoethoxydiphenyl borate and flufenamic acid inhibit Ca^{2+} influx through TRPM2 channels in rat dorsal root ganglion neurons activated by ADP-ribose and rotenone. *J. Membr. Biol.* **241,** 69–75 (2011).
24. Usoskin, D. *et al.* Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. *Nature Neurosci.* **18,** 145–153 (2015).
25. Hara, Y. *et al.* LTRPC2 Ca^{2+}-permeable channel activated by changes in redox status confers susceptibility to cell death. *Mol. Cell* **9,** 163–173 (2002).
26. Togashi, K., Inada, H. & Tominaga, M. Inhibition of the transient receptor potential cation channel TRPM2 by 2-aminoethoxydiphenyl borate (2-APB). *Br. J. Pharmacol.* **153,** 1324–1330 (2008).
27. Yamamoto, S. *et al.* TRPM2-mediated Ca^{2+} influx induces chemokine production in monocytes that aggravates inflammatory neutrophil infiltration. *Nature Med.* **14,** 738–747 (2008).
28. Uchida, K. *et al.* Lack of TRPM2 impaired insulin secretion and glucose metabolisms in mice. *Diabetes* **60,** 119–126 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Tolkovsky for help and advice, R.-L. Yu for assistance with experiments, Y. Mori for *Trpm2*$^{-/-}$ mice and for the cDNA for mouse TRPM2, S. Birren for MAH cells, J. Wood for loan of the thermal-preference test apparatus, S. Skerratt for PF-4674114, and E. Smith, J. Btesh, D. Andersson, and T. Buijs for their comments on the manuscript. Initial experiments were performed in the Department of Pharmacology, University of Cambridge. Supported by a grant from the BBSRC (UK) to P.A.McN. and a Raymond & Beverley Sackler studentship to C.-H.T.

Author Contributions C.-H.T. designed and performed experiments, analysed data and wrote the paper. P.A.M. designed experiments, analysed data, supervised the work and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.A.M. (peter.mcnaughton@kcl.ac.uk).

Reviewer Information *Nature* thanks M. Caterina, Y. Mori and the other anonymous reviewer(s) for their contribution to the peer review of this work.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

Animals. All *in vitro* experiments used C57BL/6 mice younger than 5 weeks old, except for adult DRG temperature threshold experiments (for example, Fig. 1c, d), in which 3-month-old adult mice were used (these mice were from the same group as was used for two-plate thermal preference tests experiments, see Fig. 3). *Trpm2*$^{-/-}$ mice were gifts from Y. Mori, and were generated as reported previously[27,28]. Mice were maintained on a 12 h day/12 h night cycle. All mice used in two-plate thermal preference tests had been backcrossed onto the parental C57Bl6/6J strain for 7 generations, and wild-type and *Trpm2*$^{-/-}$ mice were littermates from breeding pairs of *Trpm2*$^{+/-}$ heterozygote mice.

Primary neuron cultures. PPG, SCG and paravertebral chain ganglia were extracted from 3 or more mice and DRG from a single mouse. Ganglia were incubated in papain (2 mg ml^{-1} in Ca^{2+}-free and Mg^{2+}-free HBSS) for 30 min at 30 °C, followed by incubation in collagenase (2.5 mg ml^{-1} in Ca^{2+}-free and Mg^{2+}-free HBSS) for 30 min at 37 °C. Ganglia were re-suspended and mechanically dissociated in Neurobasal-A/B27 growing medium, which was prepared with Neurobasal-A Medium supplemented with 0.25% (v/v) L-glutamine 200 mM (Invitrogen), 2% (v/v) B-27 supplement (Invitrogen), 1% (v/v) penicillin-streptomycin (Invitrogen), and nerve growth factor (NGF) (Sigma-Aldrich) at 50 ng ml^{-1}. Dissociated neurons were centrifuged and plated onto coverslips pre-coated with poly- L-lysine (10 μg ml^{-1}) and laminin (40 μg ml^{-1}). Neurons were kept in a 37 °C incubator with a 95% air / 5% CO_2 atmosphere for at least 3 h before use, and all neurons were used within 24 h.

PC12 cell cultures. The growth medium used for PC12 cell culture was RPMI-1640 (Sigma-Aldrich), supplemented with: 1% (v/v) penicillin-streptomycin (Invitrogen), 1% (v/v) L-glutamine 200 mM (Invitrogen), 10% (v/v) horse serum (Invitrogen) and 5% (v/v) fetal bovine serum (FBS, Invitrogen). The differentiation medium for PC12 cells was RPMI-1640 supplemented with: 1% (v/v) penicillin-streptomycin (Invitrogen), 1% (v/v) L-glutamine-200 mM (Invitrogen), 1% (v/v) horse serum (Invitrogen), and NGF (Sigma-Aldrich) with final concentration at 100 ng ml^{-1}. PC12 cells were incubated and maintained in a 37 °C incubator with a 95% air / 5% CO_2 atmosphere. Medium was changed every 2 days, and cells were split every 3–4 days when grown to 90% confluency. The PC12 cells were seeded for imaging on coverslips pre-coated with poly-L-lysine (1 mg ml^{-1}; Sigma-Aldrich) and collagen IV (1 mg ml^{-1}; Sigma-Aldrich). PC12 cell lines were not authenticated and were not tested for mycoplasma contamination.

MAH cell cultures. MAH cells were kind gifts from A. Tolkovsky and S. Birren[18]. The growth medium used for MAH cell culture was L-15 medium (Sigma-Aldrich) supplemented with: 1% (v/v) penicillin-streptomycin (Invitrogen), 1% (v/v) L-glutamine 200 mM (Invitrogen), 10% (v/v) fetal bovine serum (FBS, Invitrogen), 17% (v/v) $NaHCO_3$ (150 mM), and dexamethasone (Sigma-Aldrich) at 5 μM. The differentiation medium for MAH cells was the same as the growth medium except dexamethasone was replaced with a cocktail of neurotrophic factors: CNTF (10 ng ml^{-1}; Peprotech), bFGF (10 ng ml^{-1}; Peprotech), and NGF (50 ng ml^{-1}; Sigma-Aldrich). Medium was changed every 2 days and MAH cells were split every 4 days when grown to 90% confluency and incubated and maintained in a 37 °C incubator with a 95% air / 5% CO_2 atmosphere. MAH cells used for imaging were seeded on coverslips pre-coated with poly-L-lysine (1 mg ml^{-1}; Sigma-Aldrich) and laminin (40 μg ml^{-1}; BD Science). MAH cell lines were not authenticated and were not tested for mycoplasma contamination.

Extracellular solutions and perfusion system for electrophysiology. Unless otherwise specified, all experiments were carried out with extracellular solution containing 140 mM NaCl, 4 mM KCl, 1.8 mM $CaCl_2$, 1 mM $MgCl_2$, 10 mM HEPES and 5 mM glucose; pH was adjusted to 7.4 with NaOH and osmolarity was between 295–305 mOsm. Sodium-free extracellular solution is prepared with the formulation above except for replacing sodium chloride with equimolar choline chloride. Calcium-free extracellular solution is prepared with the formulation above except for removal of calcium chloride. An 8-line manifold gravity-driven system controlled by an automated solution changer with a common outlet was used to apply solution to the cells. The temperature in three lines was heated or cooled with a Peltier device regulated by a proportional gain feedback controller designed by V. Vellani (CV Scientific). The temperature in each experimental protocol was recorded by a miniature thermocouple immediately before the solution entered the bath or (in separate control experiments) at the cell location at the tip of the solution outlet. All compounds applied were prepared as stock solutions first and then diluted to the concentration needed before experiments. Capsaicin was dissolved in ethanol to make 5 mM stock solutions. Pregnenolone sulphate was dissolved in DMSO to make 500 mM stock solutions. 2-APB was dissolved in DMSO to make 500 mM stock solutions. The TRPV4 agonist, PF-4674114, was dissolved in DMSO to make 5 mM stock solutions. Nifedipine was dissolved in DMSO to make 100 mM stock solutions. TTX was dissolved in pH 4.8 citrate buffer to make 100 mM stock solutions. Verapamil, ruthenium red, and H_2O_2 were dissolved in extracellular solution on the day of experiments.

Calcium imaging. Cells were loaded with 5 μM fura-2 AM (Invitrogen) with 0.02% (v/v) pluronic acid (Invitrogen) for 30 min. After loading, coverslips were put in an imaging chamber and transferred to a Nikon Eclipse Ti-E inverted microscope. Cells were continuously perfused with extracellular solution and were illuminated with a monochromator alternating between 340 and 380 nm (OptoScan; Cairn Research), controlled by WinFluor 3.2 software (J. Dempster, University of Strathclyde, UK). Emission was collected at 510 nm and the resulting pairs of images were acquired every two seconds with a 100 ms exposure time using an iXon 897 EM-CCD camera (Andor Technology, Belfast, UK). Image time series were converted to TIFF files and processed with ImageJ software. Images of the background fluorescence intensity were obtained for both wavelengths and subtracted from the respective image stack before calculating the $F_{340/380}$ ratio images. A minority of neurons (<10%) exhibited an unstable $F_{340/380}$ baseline in the absence of any applied stimulus, usually caused by poor dye loading but in some cases apparently due to low-frequency repetitive firing even in the absence of any treatment, and were removed from analysis. In experiments to identify neurons responding to known TRP channel agonists (Fig. 1a), we found that PS caused a very slow increase in $F_{340/380}$ ratio in some neurons, clearly distinguishable from the rapid elevation in $[Ca]_i$ seen in TRPM3-expressing DRG neurons. This slow response can probably be attributed to an off-target effect of PS as it was also seen in autonomic neurons which do not appear to express TRPM3 (Fig. 2). A positive response to all agonists was therefore defined from the rate of increase of $[Ca]_i$ following agonist application, as an increase of $F_{340/380}$ ratio, between two consecutive time points following application of agonist, which exceeds the mean + 3.09 s.d. (cumulative probability value of 99.9%) of all such differences in the absence of any agonist. A heat-sensitive neuron is defined as a neuron with a peak increase in $F_{340/380}$ during a heat stimulus larger than the mean + 3.09 s.d. of the peak increase in $F_{340/380}$ of the glial cells in the same experiment (see Extended Data Fig. 1b). The thermal threshold of a heat-responsive neuron (see Fig. 1d, e) was defined as the temperature when the increase in $F_{340/380}$ ratio between two consecutive time points is larger than the mean + 3.09 s.d. of the increase in $F_{340/380}$ of the glial cells between two consecutive time points in the same experiment. For MAH and PC12 cell cultures, where no glial cells were present, we used the value of mean + 3.09 s.d. obtained from glial cells in similar experiments on neuronal cultures.

Supplementary Information Video 1 shows an example series of calcium images.

Patch clamp recordings. The intracellular solution for concurrent calcium imaging and patch clamp (see Fig. 2b) contained 140 mM KCl, 1.6 mM $MgCl_2$, 2.5 mM MgATP, 0.5 mM NaGTP, 10 mM HEPES and 167 μM fura-2; pH was adjusted to 7.3 with KOH. The intracellular solution for current-voltage relationship determination, in which Ca^{2+} and K^+ currents were blocked (see Fig. 2c), contained 130 mM CsCl, 2.5 mM MgATP, 0.5 mM NaGTP, 10 mM HEPES, 10 mM TEA, and 5 mM 4-AP; pH was adjusted to 7.3 with CsOH. The osmolarity of both of the intracellular solutions was between 295–305 mOsm. The extracellular solution for current–voltage relationship determination contained 125 mM NaCl, 2 mM $CaCl_2$, 10 mM HEPES and 5 mM glucose, 10 mM TEA, 5 mM 4-AP, 2 μM tetrodotoxin, and 100 μM $CdCl_2$.

All patch clamp experiments were carried out with an Axopatch 200B patch-clamp amplifier (Axon Instrument, USA). Patch pipettes (Blaubrand 100 μl borosilicate glass, Scientific Laboratory Supplies, Germany) were pulled using a Flaming/Brown P-97 horizontal micropipette puller (Sutter Instruments, USA) and had a resistance between 3 and 5.5 MΩ. A giga-ohm seal was formed between the patch pipette and the cell membrane and the pipette capacitance transients were cancelled before achieving the whole cell configuration. All experiments were begun in voltage clamp mode with holding potential at −60 mV at the time of entering whole cell mode. After entering the whole-cell mode, series resistance was adjusted to be lower than 20 mega-ohm. Resting membrane potential was tested and only neurons with membrane potentials more negative than −50 mV were used for recording. Data were acquired and analysed with pClamp10 software (Axon Instruments, USA) and whole cell currents and voltages were filtered at 1 kHz and sampled at 10 kHz.

RNA-seq. MAH cells were trypsinized and collected as a cell pellet before lysis. RNA extraction was performed with the miRNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. Two samples from MAH cells grown in growth medium and 2 samples from MAH cells grown in differentiation medium were sent to Oxford Gene Technology to complete the rest of the steps for RNA-sequencing. Sequencing libraries were prepared with the Illumina TruSeq RNA Sample Prep Kit v2. A total of 4 samples (two cold-sensitive MAH cells and two cold-insensitive MAH cells) were sequenced on 2 lanes on the Illumina HiSeq2000 platform using TruSeq v3 chemistry. All sequences were paired-end and sequencing was performed over 100 cycles. Read files (Fastq) were generated from the sequencing

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

platform via the manufacturer's proprietary software. Reads were processed through the Tuxedo suite. Reads were mapped to their location to the appropriate Illumina iGenomes build using Bowtie version 2.02. Splice junctions were identified using TopHat, version v2.0.9. Cufflinks version 2.1.1 was used to perform transcript assembly, abundance estimation and differential expression and regulation for the samples. Visualization of differential expression results were performed with CummeRbund. RNA-seq alignment metrics were generated with Picard.

***In situ* hybridization following calcium imaging.** Coverslips were marked on the periphery with a diamond knife to assist localization of the imaged region and were then calcium-imaged as above to identify novel heat-sensitive neurons. Following calcium imaging, coverslips were rinsed with phosphate-buffered saline (PBS) then fixed in 4% paraformaldehyde (PFA) at 4 °C for 20 min. TRPM2 mRNA was detected with digoxigenin-labelled antisense probes against mouse *Trpm2* (NM_138301.2). We are very grateful to Y. Mori of Kyoto University for providing the mouse *Trpm2* gene cloned into the pCI-neo plasmid (Promega).

Probe synthesis. For the *Trpm2* antisense probe the plasmid was linearized with EcoRI and transcribed with T3 RNA polymerase and for the *Trpm2* sense probe, which was used as a negative control, the plasmid was linearized with SalI and transcribed with T7 RNA polymerase.

Hybridization. Fixed coverslips were rinsed in PBS with 0.1% Triton, and were then incubated in *in situ* hybridization solution without probe at 47 °C for 30 min as pre-hybridization step. After pre-hybridization, coverslips were transferred into *in situ* hybridization solution with antisense or sense probes for hybridization at 47 °C overnight[29].

Post-hybridization steps. Following hybridization coverslips were washed in 2× SSC and then 0.2× SSC at 47 °C for 30 min for each solution. Coverslips were then washed twice with KTBT solution at room temperature for 5 min for each washing. 25% normal goat serum was then used for blocking cells for 1 h at room temperature. Coverslips were then incubated in 25% normal goat serum containing pre-absorbed anti-digoxygenin antibody coupled to alkaline phosphatase for 2 h at room temperature, followed by washing 3× in KTBT for 15 min each wash, and then twice in alkaline phosphatase buffer at room temperature for 10 min each wash. Coverslips were then developed in alkaline phosphatase buffer containing 337.5 μg ml^{-1} NBT and 175 μg ml^{-1} BCIP in the dark for 8 h before being washed in KTBT, fixed in 4% PFA for 10 min, washed in PBS, and then mounted in SlowFade Gold Antifade Mountant with DAPI[29]. DIC transmitted-light images were acquired through a Plan Fluor 10× Ph1 DLL objective with a DS-Qi2 monochrome camera on a Nikon Eclipse Ti-E inverted microscope. A GFPHQ filter was used to enhance the dark purple colour. The images were rotated, cropped, and resized with ImageJ to be aligned with the images obtained in calcium imaging.

Staining for isolectin B4 (IB4) following calcium imaging. DRG neurons on marked coverslips were calcium imaged as above then rinsed with phosphate-buffered saline (PBS) and fixed in 4% paraformaldehyde (PFA) at 4 °C for 20 min. After fixation, coverslips were washed twice in PBS with 0.1% Triton then incubated in solution containing 10 μg ml^{-1} IB4 bound to Alexa Fluor 594, 10% normal goat serum, 2% bovine serum albumin, 0.1% Triton, and 10 mM sodium azide for 1 h at room temperature followed by washing with PBS three times. DIC transmitted-light images were acquired through a Plan Fluor 10× Ph1 DLL objective with a DS-Qi2 monochrome camera on a Nikon Eclipse Ti-E inverted microscope. A Texas Red HYQ filter was used to capture the Alexa 594 signal. The images obtained were rotated, cropped, and resized with ImageJ to be aligned with the images obtained in calcium imaging.

Two-plate thermal preference tests. To eliminate as far as possible any extraneous genetic influences the *Trpm2*$^{-/-}$ mice were backcrossed onto the parental C57Bl6/6J line for 7 generations[27,28]. To minimize environmental effects, wild-type and *Trpm2*$^{-/-}$ littermates from heterozygous matings were compared in behavioural experiments. Sample size to achieve significance was determined from trial experiments but no power analysis was performed. All mice were tested at all temperatures so no randomization of experimental groups was necessary (see Fig. 3 legend). We used a two-plate thermal preference test (BioSeb, France) with one plate maintained at a temperature of 33 °C, which other studies have shown is the preferred temperature[11], and the other at a variable temperature. The temperatures of test and control plates were reversed after 30 min to control for any influence of environmental cues. Other studies have observed sex differences in mouse thermal behaviour[12] so we followed other authors[5,11] in using only adult males (10–16 weeks old) in behavioural experiments.

Two hot/cold-plate machines (Bioseb, France), placed back to back, formed the two-plate thermal apparatus. Plates were enclosed in a plexiglass chamber divided into two lanes, with an opaque compartment between them, and two mice were tested simultaneously in adjacent lanes (see Supplementary Information Video 2). The temperature of each plate was controlled by T2CT software (Bioseb, France). Plate temperatures were tested with an infrared thermometer (Bioseb) and were found to be accurately controlled to within 0.2 °C of the command temperature over the entire plate area. One plate was maintained at the preferred temperature of 33 °C, and mice were initially placed onto the plate with starting temperature other than 33 °C ('plate A', see Fig. 3) before initiating recording. The movements of the mice between the two plates were recorded for 3,495 s without human presence and the mouse position was determined with an automated video tracking system (Bioseb), so operator blinding was not necessary. The temperatures of the two plates were exchanged 1,800 s after initiation of recording; plate temperature settled to within 0.2 °C of the new temperature within 180 s of the change. Experiments were performed between 8 a.m. and 10 p.m., with room temperature at 20 °C. For experiments testing thermal preference between the two mildest temperatures, 28 °C versus 33 °C and 33 °C versus 38 °C, mice were tested again, with the starting temperatures of the two plates exchanged, 3–5 h after the first recording. For other temperatures recordings were made only once on a particular mouse. Mice were tested with the protocols, in order, of 28 °C versus 33 °C, 33 °C versus 38 °C, 23 °C versus 33 °C and 33 °C versus 43 °C. Sample size was based on pilot experiments. When making statistical comparisons variances were checked to ensure that it was similar between groups being compared. All animal experiments were approved by the Animal Welfare and Ethical Review Body (AWERB), King's College London.

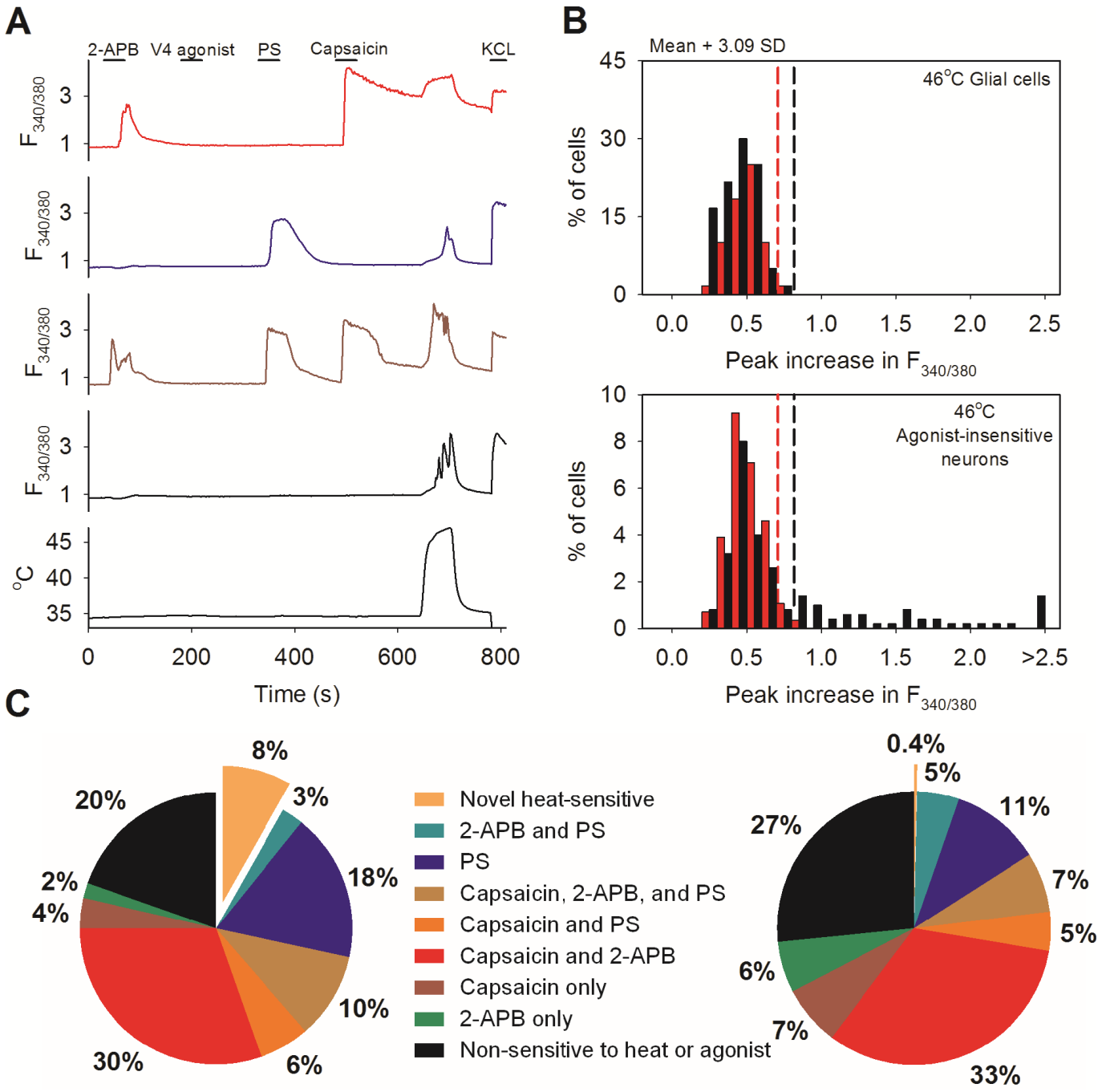
Supplementary Information Video 2 shows an example of a thermal-choice behavioural experiment.

Statistical analysis. All data are expressed as means ± s.e.m. Analyses were performed with GraphPad Prism version 6.01 or SigmaPlot 11.0. The particular statistical test used is stated either in the text or figure legends.

Biological and technical replicates. Biological replicates are stated in the legends for each figure. Given the nature of these experiments, technical replicates were not possible.

29. Ariza-McNaughton, L. & Krumlauf, R. Non-radioactive in situ hybridization: simplified procedures for use in whole-mounts of mouse and chick embryos. *Int. Rev. Neurobiol.* **47,** 239–250 (2002).
30. Schmid, D., Messlinger, K., Belmonte, C. & Fischer, M. J. Altered thermal sensitivity in neurons injured by infraorbital nerve lesion. *Neurosci. Lett.* **488,** 168–172 (2011).
31. Jang, Y. *et al.* TRPM2 mediates the lysophosphatidic acid-induced neurite retraction in the developing brain. *Pflugers Arch.* **466,** 1987–1998 (2014).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

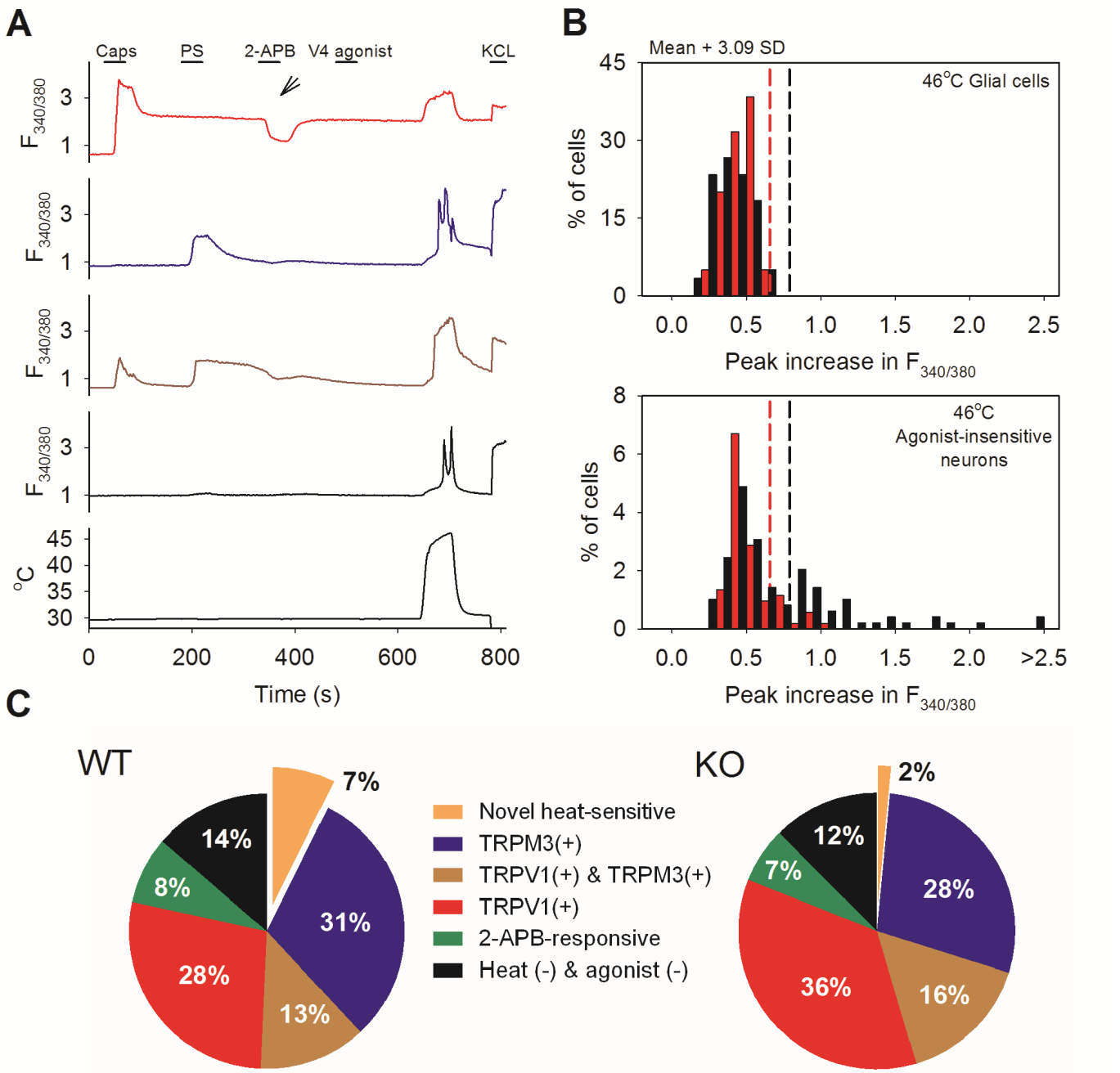


Extended Data Figure 1 | See next page for caption.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

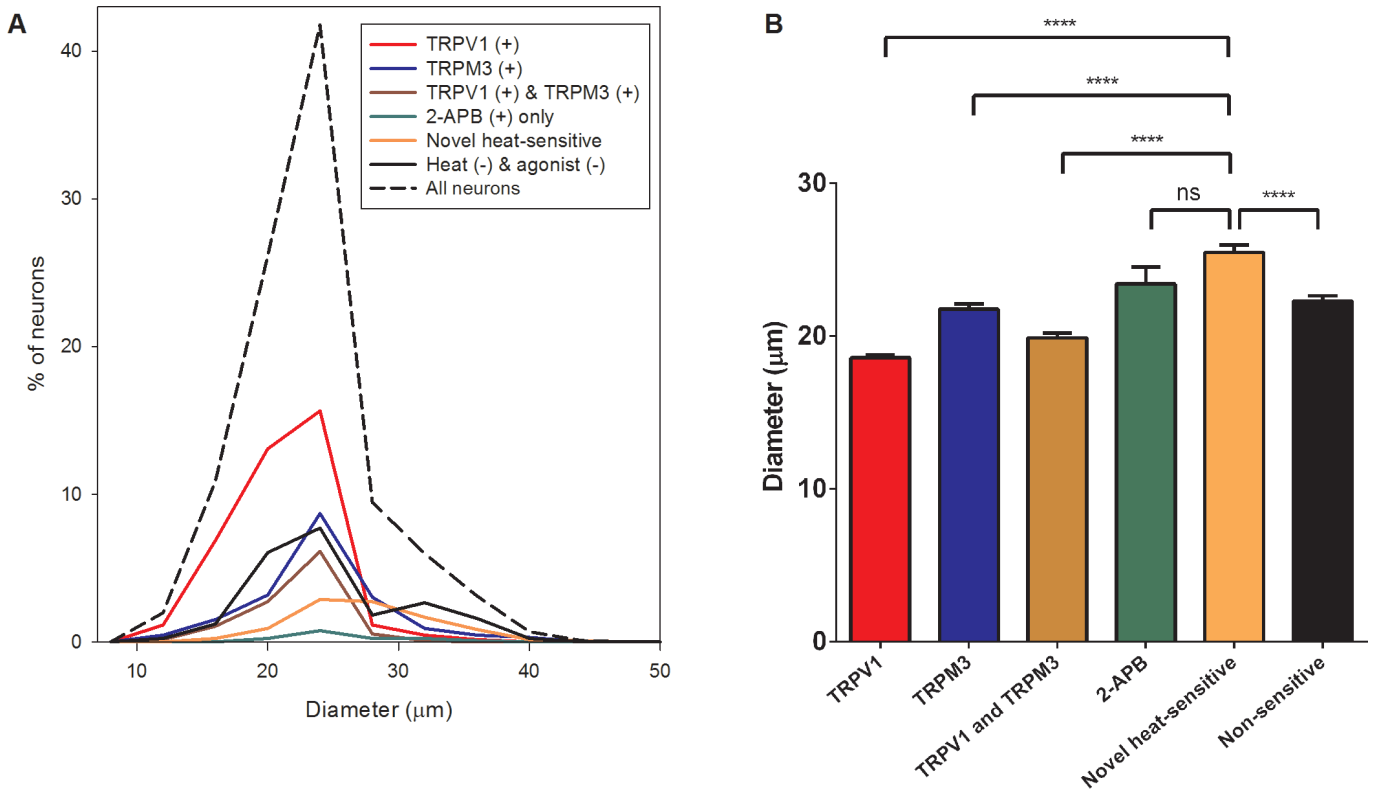
Extended Data Figure 1 | Effect of altering the order of agonist application in DRG neurons a, Method for detecting novel heat-sensitive somatosensory neurons. Representative traces showing increases of $[Ca^{2+}]_i$ ($F_{340/380}$ ratio, ordinate) in DRG neurons in response to the TRPV1-3 agonist 2-APB (250 μM), to the specific TRPV4 agonist PF-4674114 (V4 agonist, 200 nM), to the specific TRPV1 agonist capsaicin (1 μM), to the TRPM3 agonist pregnenolone sulphate (PS, 100 μM), to a heat ramp from 35 °C to 46 °C (temperature protocol shown at bottom), and to KCl (50 mM). Other details as in Fig. 1. From top: TRPV1-expressing neuron responding to 2-APB, capsaicin (1 μM) and heat (red, 30% of 500 neurons); TRPM3-expressing neuron responding to PS (blue, 100 μM) and heat (18%); TRPV1-and TRPM3-co-expressing neuron responding to 2-APB, PS, capsaicin, and heat (brown, 10%); neuron unresponsive to TRP channel agonists but showing a response to heat and therefore expressing a novel heat sensitive ion channel (black, 8% of total). No neuron responded to the specific TRPV4 agonist PF-4674114 (200 nM). **b**, Heat has a small effect on the fura-2 fluorescence ratio[30], so we eliminated neurons in which an increase of fluorescence ratio was due simply to this physical effect by comparing the increase of fura-2 fluorescence ratio in neurons with that in glial cells in the same culture. Maximum increases in $F_{340/380}$ ratio in response to a heat ramp from 35 °C to 46 °C in wild-type glial cells (black bars, top, $n = 60$) and in wild-type DRG neurons not responding to known thermo-TRP agonists (black bars, bottom, from same images as the glial cells in top panel, $n = 139$). Vertical black dashed line in top panel shows mean + 3.09 s.d. (cumulative probability value of 99.9%) of the increase in the $F_{340/380}$ ratio in glial cells; this value is taken as the maximum increase in $F_{340/380}$ ratio caused by effect of heat on fura-2 and is used as the cut-off value for defining novel heat-sensitive neurons present in the same culture dish (vertical black dashed line in lower panel). Similar results from separate culture of *Trpm2*$^{-/-}$ glia ($n = 40$) and neurons ($n = 76$) shown in red. The proportion of novel heat-sensitive neurons was significantly reduced from 8% (41/500) in wild type to 0.4% (1/282) in *Trpm2*$^{-/-}$ ($P \leq 0.0001$; Fisher's exact test). The increases in $F_{340/380}$ ratio of novel heat-sensitive neurons above the cut-off value in response to heat (smallest increase = 0.019854) are all higher than that of the single heat-responding *Trpm2*$^{-/-}$ neuron (0.019084). **c**, Pie charts showing the percentage of novel heat-sensitive neurons responding to TRP ion channel agonists and to heat in wild-type DRG neurons (left) and DRG neurons from *Trpm2*$^{-/-}$ mice (right). Deletion of *Trpm2* reduces the percentage of novel heat-sensitive neurons from 8% to 0.4%. Cell numbers for **a**–**c** were 500 DRG neurons from one wild-type mouse on 3 coverslips and 282 DRG neurons from one *Trpm2*$^{-/-}$ mouse on 2 coverslips imaged. No further replicates of this particular experimental protocol were performed.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



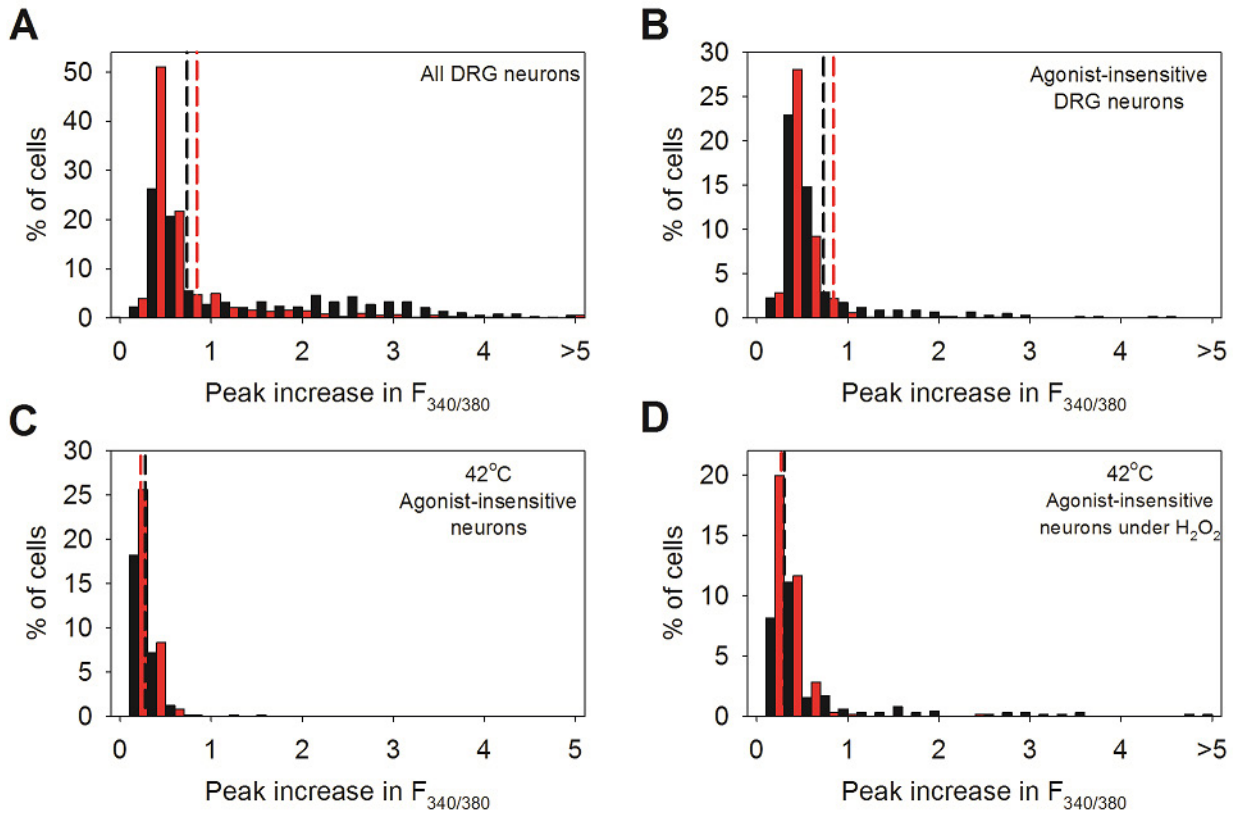
Extended Data Figure 2 | Effect of starting the heat ramp at a lower temperature in DRG neurons. Identical experiment to that shown in Fig. 1a–c, except that the temperature ramp started from 30 °C. **a**, Agonist and heat-responsive neurons as in Fig. 1a. From top: TRPV1-expressing neuron responding to capsaicin (1 μM) and heat (red, 28% of 491 neurons); TRPM3-expressing neuron responding to PS (blue, 100 μM) and heat (31%); TRPV1-and TRPM3-co-expressing neuron responding to capsaicin, PS and heat (brown, 13%); neuron unresponsive to TRP channel agonists but showing a response to heat and therefore expressing a novel heat-sensitive ion channel (black, 7% of total). A small number of neurons (8%) responded to 2-APB (250 μM) but not to other agonists, and 14% of DRG neurons did not respond to any of the agonists nor to heat (not shown). No neuron responded to the specific TRPV4 agonist PF-4674114 (200 nM). **b**, Maximum increases in $F_{340/380}$ ratio in response to a heat ramp from 30 °C to 46 °C in wild-type glial cells (black bars, top, $n = 60$) and in heat-sensitive wild-type DRG neurons not responding to known thermo-TRP agonists (black bars, bottom, from same images as the glial cells in top panel, $n = 103$). Vertical black dashed line in top panel shows mean + 3.09 s.d. (cumulative probability value of 99.9%) of the increase in the $F_{340/380}$ ratio in glial cells; this value is taken as the maximum increase in $F_{340/380}$ ratio caused by effect of heat on fura-2 and is used as the cut-off value for defining novel heat-sensitive neurons present in the same culture dish (vertical black dashed line in lower panel). Similar results from separate culture of *Trpm2*$^{-/-}$ glia ($n = 60$) and neurons ($n = 73$) shown in red. The proportion of novel heat-sensitive neurons was significantly reduced from 7% (103/491) in wild type, to 2% (73/522) in *Trpm2*$^{-/-}$ ($P \leq 0.0001$; Fisher's exact test). The mean increase in $F_{340/380}$ ratio of novel heat-sensitive neurons above the cut-off values in response to heat was also significantly reduced from 1.237 ± 0.09207 in wild type ($n = 36$) to 0.7959 ± 0.03767 in *Trpm2*$^{-/-}$ ($n = 8$) ($P = 0.0313$; two-tailed unpaired *t*-test). **c**, Pie charts showing the percentage of novel heat-sensitive and TRPV1- or TRPM3-expressing neurons in wild-type DRG neurons (left) and DRG neurons from *Trpm2*$^{-/-}$ mice (right). Cell numbers for imaging in **a**–**c** were 491 DRG neurons from one wild-type mouse on 3 coverslips and 522 DRG neurons from one *Trpm2*$^{-/-}$ mouse on 3 coverslips. No further replicates of this particular experimental protocol were performed.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



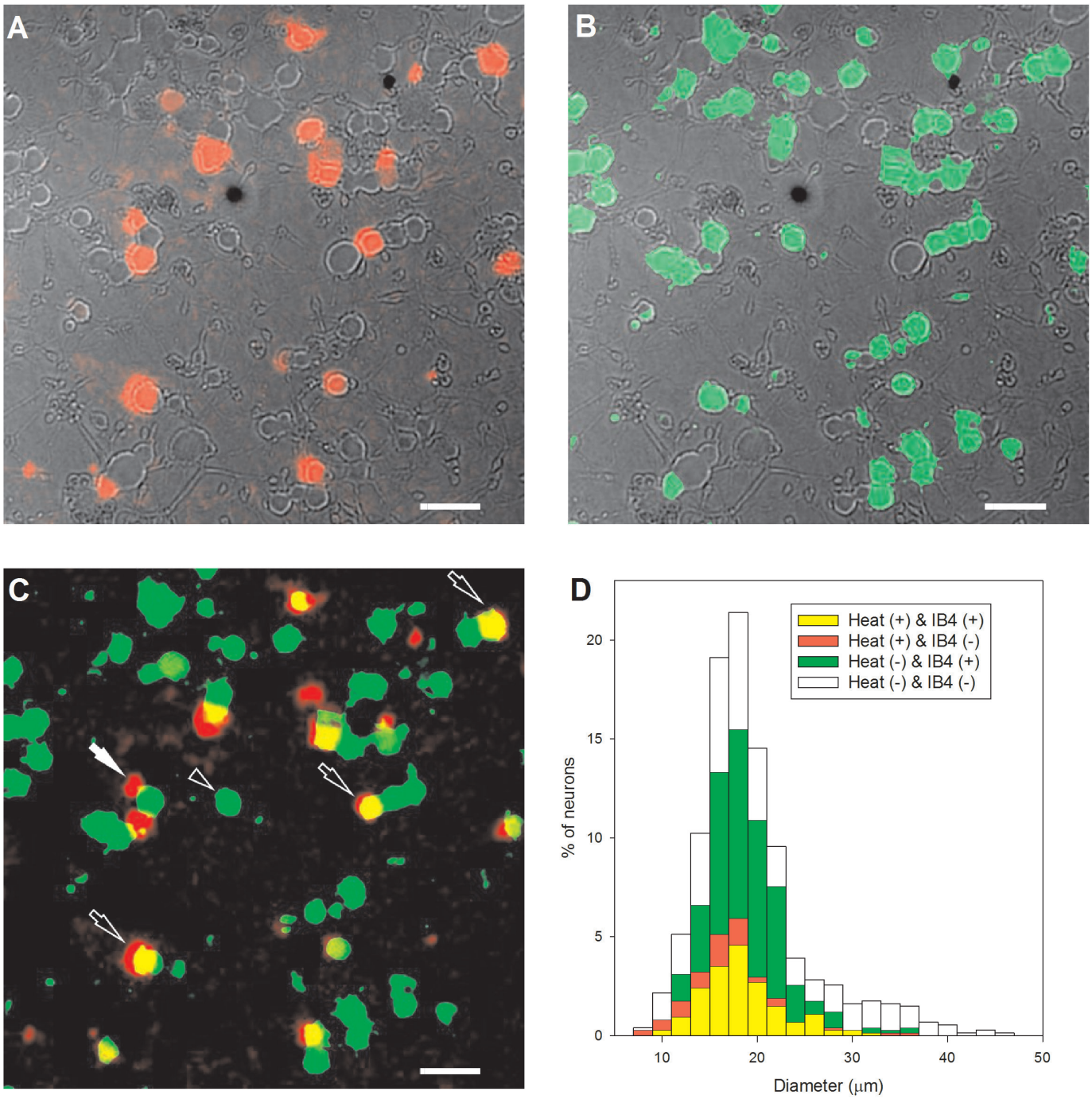
Extended Data Figure 3 | Diameters of novel heat-sensitive DRG neurons compared to neurons responding to other TRP agonists. **a**, Diameters of 1,324 DRG neurons taken from experiment illustrated in Fig. 1a (dotted line). Subpopulations of neurons are shown as follows: those responding to capsaicin and thus expressing TRPV1 (red); to PS and thus expressing TRPM3 (blue); to both agonists and thus co-expressing TRPV1 and TRPM3 (brown); to 2-APB alone (green); novel heat-sensitive neurons (orange), and neurons responding neither to heat nor to any of these agonists (black). **b**, Diameter comparison of subpopulations of neurons. TRPV1-expressing neurons have the smallest mean diameter ($18.58 \pm 0.17\,\mu m$), TRPM3-expressing neurons are intermediate ($21.75 \pm 0.33\,\mu m$), and neurons expressing only the novel heat-sensitivity have the largest mean diameter ($25.47 \pm 0.48\,\mu m$). Significance from one-way ANOVA and multiple comparisons with Tukey's multiple comparison test (****$P \leq 0.0001$; NS, not significant). Data obtained from Fig. 1a–c; 1,324 DRG neurons from one wild-type mouse on 4 coverslips were analysed.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



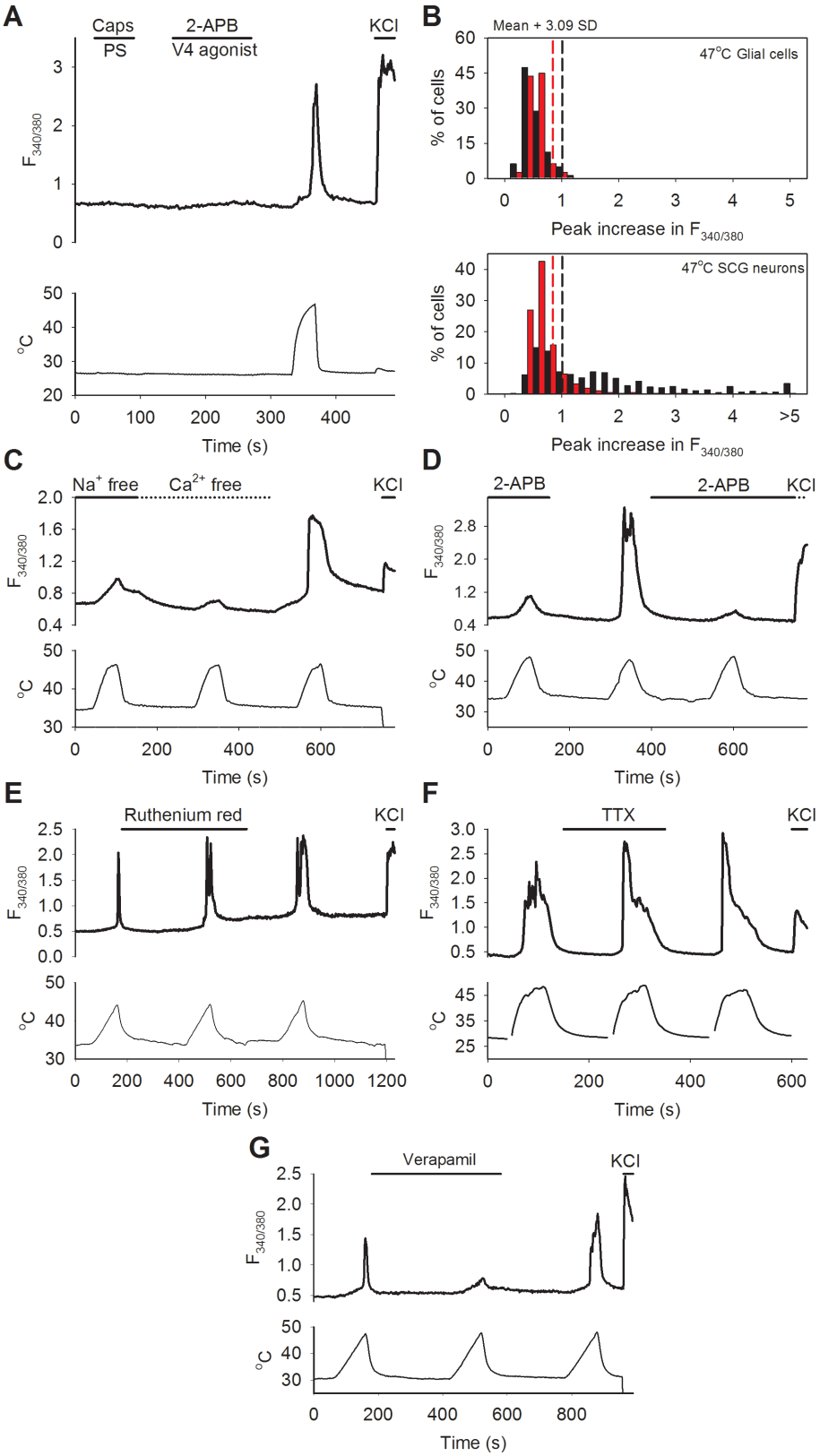
Extended Data Figure 4 | The novel heat-sensitivity in DRG neurons is partially co-expressed with TRPV1 and TRPM3, and is enhanced by H_2O_2. **a**, Temperature ramp to 47 °C, as in Fig. 1a, but with TRPV1 blocked with AMG9810 (5 μM) and TRPM3 blocked with naringenin (10 μM). Criterion level for significant increase (dashed lines) taken from glial cells in same field of view (data not shown). Black bars: 46% of all wild-type DRG neurons ($n = 580$) responded to heat ramp from 34 °C to 46 °C with an increase in $[Ca^{2+}]_i$ above criterion level in presence of blockers of TRPV1 and TRPM3 (dashed vertical line), while the percentage decreased to 17% in *Trpm2*$^{-/-}$ DRG neurons (red bars, $n = 1{,}007$) ($P \leq 0.0001$; Fisher's exact test). The mean increase in $F_{340/380}$ ratio above the cut-off values (dashed lines) in response to heat was also significantly reduced from 1.619 ± 0.06133 in wild type ($n = 265$) to 1.027 ± 0.08394 in *Trpm2*$^{-/-}$ ($n = 175$) ($P \leq 0.0001$; two-tailed unpaired *t*-test). In similar experiments with TRPV1 blocker BCTC (4 μM) and naringenin (10 μM), 37% of wild-type neurons responded to heat (data not shown, $n = 554$). **b**, Similar plot as in **a**, but data from the subgroup of novel heat-sensitive neurons not responding to agonists for known thermo-TRP channels. After exposure to heat in presence of blockers of TRPV1 and TRPM3, blockers were removed and neurons not responding to known TRP agonists were identified as in Fig. 1a. Data from same experiment as shown in **a**. Proportion of neurons expressing the novel heat-sensitive mechanism in isolation (that is, without co-expression of TRPV1 or TRPM3) was significantly lower (52/580, 9%) than all neurons expressing the novel heat-sensitive mechanism (46%, see **a**). The proportion of novel heat-sensitive neurons was significantly reduced in *Trpm2*$^{-/-}$ mice, from 9% to 0.6% (6/1,007, $P \leq 0.0001$; Fisher's exact test). **c**, Temperature ramp to 42 °C. Few novel heat-sensitive neurons respond to this low temperature in either wild type or *Trpm2*$^{-/-}$. The total number of neurons was $n = 173$ for wild type and $n = 211$ for *Trpm2*$^{-/-}$. **d**, Responses of the same neurons to the same temperature ramp to 42 °C following addition of H_2O_2 (400 μM). Enhancement of response in wild type (black bars) was largely abolished in *Trpm2*$^{-/-}$ (red bars). The proportion of novel heat-sensitive neurons after sensitization with H_2O_2 was significantly reduced from 11% (74/635) in wild type to 8% (48/601) in *Trpm2*$^{-/-}$ ($P = 0.0356$; Fisher's exact test). The mean increase in $F_{340/380}$ ratio of novel heat-sensitive neurons above the cut-off values in response to heat was also significantly reduced from 1.175 ± 0.1516 in wild type ($n = 72$) to 0.4485 ± 0.04329 in *Trpm2*$^{-/-}$ ($n = 48$) ($P = 0.0002$; two-tailed unpaired *t*-test). Cell numbers and replicates for **a** and **b** were 580 DRG neurons from one wild-type mouse on 5 coverslips and 1,007 DRG neurons from one *Trpm2*$^{-/-}$ mouse on 5 coverslips imaged for the protocol with AMG9810. 554 neurons from one wild-type mouse on 4 coverslips were imaged for the protocol with BCTC. No further replicates were carried out. The cell numbers and replicates for **c** and **d** were 635 DRG neurons from one wild-type mouse on 3 coverslips and 601 DRG neurons from one knockout mouse on 3 coverslips imaged. The experiment was replicated with similar results on 4 additional coverslips from one mouse.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 5 | Most novel heat-sensitive DRG neurons are IB4-positive. a, Increases in $[Ca^{2+}]_i$ ($F_{340/380}$ ratio image, intensity-coded in red, in mouse DRG neurons) in response to heat ramp to 46 °C (TRPV1 blocked with AMG9810, 5 μM, and TRPM3 blocked with naringenin, 10 μM), superimposed on differential interference contrast (DIC) transmission image obtained post-fixation. **b**, The same field following fixation and labelling with fluorescent IB4 (green). **c**, Superimposed calcium and IB4 images from **a** and **b**. Black arrows show neurons responding to heat and positive for IB4. White arrow indicates a neuron responding to heat and negative for IB4. Black arrowhead shows neuron insensitive to heat and positive for IB4. Scale bars 50 μm. **d**, Diameter histogram of 743 fixed DRG neurons subgrouped according to novel heat-sensitivity (yellow, red) and IB4 binding (yellow, green). 25% (184/743) of DRG neurons showed novel heat-sensitivity, and 74% of these novel heat-sensitive neurons were IB4-positive, whereas only 53% of heat-insensitive neurons were IB4-positive. The percentage of IB4-positive neurons is significantly higher in the heat-sensitive group than in the heat-insensitive group ($P \leq 0.0001$; Fisher's exact test). The diameters shown in **d** are not directly comparable with the live cell diameters shown in Extended Data Fig. 3 because of a shrinkage artefact on fixation. Cell numbers were 743 DRG neurons from one wild-type mouse on 4 coverslips imaged. No further replicates were performed.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

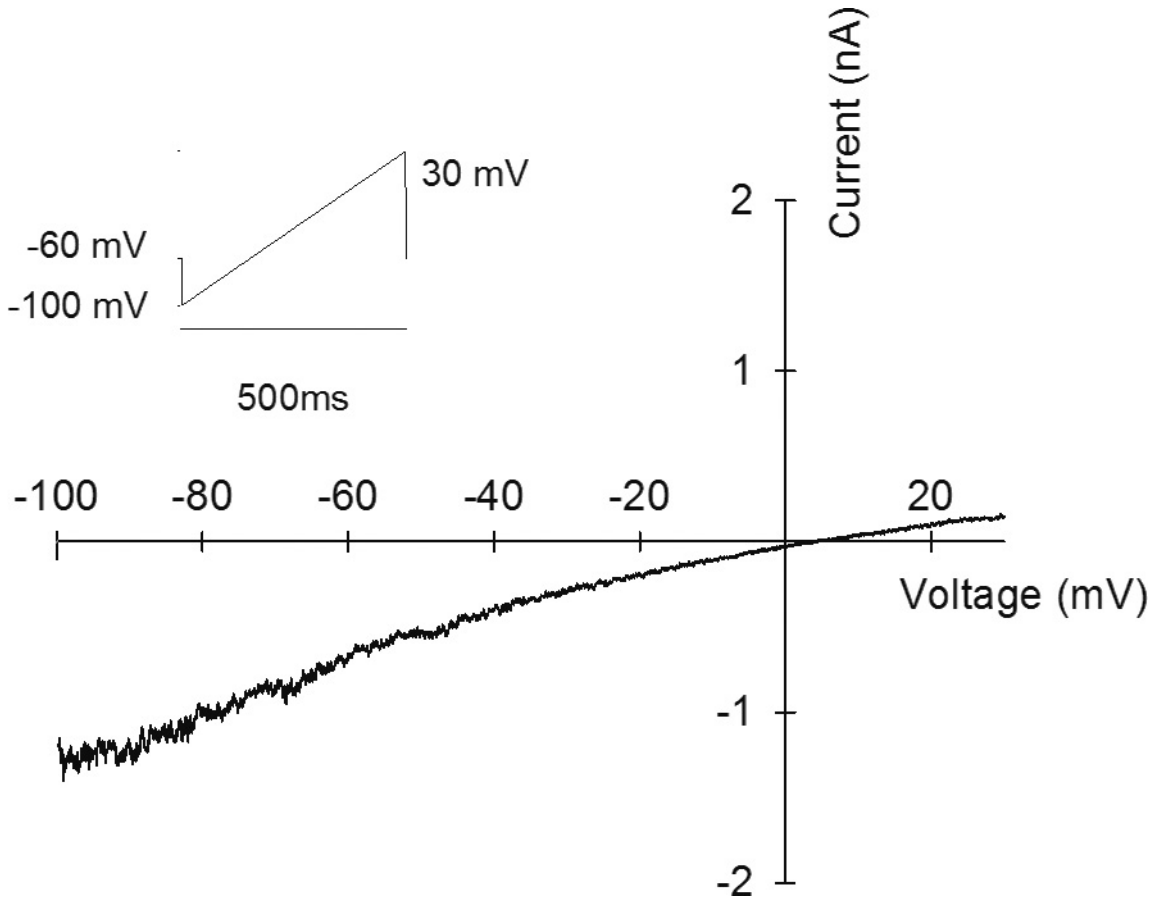


Extended Data Figure 6 | See next page for caption.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

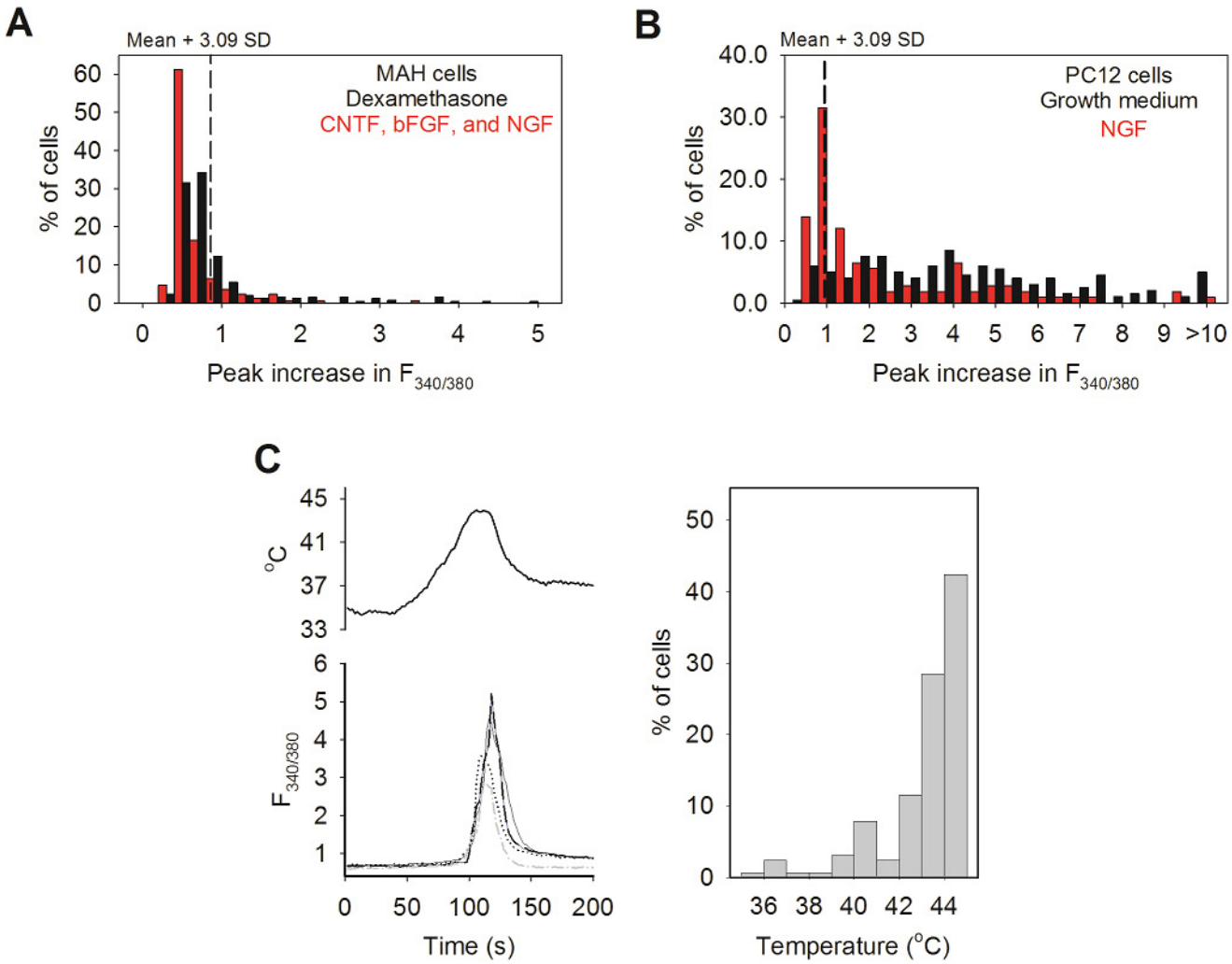
Extended Data Figure 6 | Properties of novel heat-sensitive ion channel expressed in autonomic neurons. a, Representative traces showing increases of $[Ca^{2+}]_i$ ($F_{340/380}$ ratio, ordinate) in sympathetic neurons from superior cervical ganglion (SCG) in response to a mixture of capsaicin (TRPV1 agonist,1 μM) and pregnenolone sulphate (TRPM3 agonist, 100 μM); a mixture of 2-APB (TRPV1-3 agonist, 250 μM) and PF-4674114 (TRPV4 agonist, 200 nM); heat to 47 °C (temperature protocol shown at bottom); and KCl (50 mM). Similar results were obtained with parasympathetic neurons from pterygopalatine ganglion (PPG, data not shown). Trace is same as shown in Fig. 2a. **b**, Similar histograms as in Extended Data Figs 1b and 2b, but for SCG glial cells and neurons from wild-type mice (black bars) and *Trpm2*$^{-/-}$ mice (red bars). 58% of wild-type SCG neurons ($n = 436$) showed novel heat-sensitivity with increases in $F_{340/380}$ ratio above the criterion level obtained from glial cells ($n = 80$) in same culture (black vertical dashed line). In similar experiments on PPG neurons ($n = 484$), 47% showed novel heat-sensitivity (not shown). Red bars and red dashed line show results from SCG glia ($n = 80$) and neurons ($n = 430$) from *Trpm2*$^{-/-}$ mice. The proportion of novel heat-sensitive neurons was significantly reduced by deletion of *Trpm2*, from 58% (252/436) in wild type to 12% (53/430) in *Trpm2*$^{-/-}$ ($P \leq 0.0001$; Fisher's exact test). The mean increase in $F_{340/380}$ ratio of heat-sensitive neurons above the cut-off values in response to heat was also significantly reduced, from 1.629 ± 0.1928 in wild type ($n = 252$) to 0.5050 ± 0.1270 in *Trpm2*$^{-/-}$ ($n = 53$) ($P = 0.0086$; two-tailed unpaired *t*-test). **c**, Heat-evoked Ca^{2+} increase in SCG neurons is reduced but not abolished by removal of extracellular Na^+ (replaced with choline) and is abolished by removal of extracellular Ca^{2+} (remaining small increase in $F_{340/380}$ ratio is due to temperature sensitivity of fura-2, see **b**). Similar results were seen in 133 SCG neurons. **d**, Heat-evoked Ca^{2+} increase in SCG neurons is blocked by TRPV agonist 2-APB (25 μM). Similar results were seen in 130 SCG neurons. **e**, The Ca^{2+} increase in PPG neurons is not affected by TRPV channel blocker ruthenium red (50 μM). Similar results were seen in 75 PPG neurons. **f**, The Ca^{2+} increase in PPG neurons is not affected by the Na channel blocker tetrodotoxin (2 μM). Similar results were seen in 35 PPG neurons. **g**, The Ca^{2+} influx in PPG neurons is reduced but not eliminated by the L-type Ca^{2+} channel blocker verapamil (100 μM). Similar results were seen in 30 PPG neurons. Cell numbers and replicates for **a** were 166 SCG neurons from three wild-type mice on 3 coverslips imaged. Cell numbers and replicates for **b** were 436 SCG neurons from two wild-type mice on 4 coverslips and 430 SCG neurons from two *Trpm2*$^{-/-}$ mice on 4 coverslips imaged. Similar results as those shown for wild-type obtained with 15 further coverslips of SCG neurons from 9 wild-type mice and 7 coverslips of PPG neurons from 6 wild-type mice. Cell numbers and replicates for **c** were 133 SCG neurons from 3 wild-type mice on 5 coverslips that showed similar responses. Cell numbers and replicates for **d** were 130 SCG neurons from 3 wild-type mice on 2 coverslips hat showed similar responses. Similar results were also obtained for DRG neurons (4 coverslips from 1 mouse). Cell numbers and replicates for **e** were 75 PPG neurons from 3 wild-type mice on 4 coverslips that showed similar responses. Cell numbers and replicates for **f** were 35 PPG neurons from 3 wild-type mice on 4 coverslips that showed similar responses. Cell numbers and replicates for **g** were 30 PPG neurons from 3 wild-type mice on 2 coverslips that showed similar responses.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



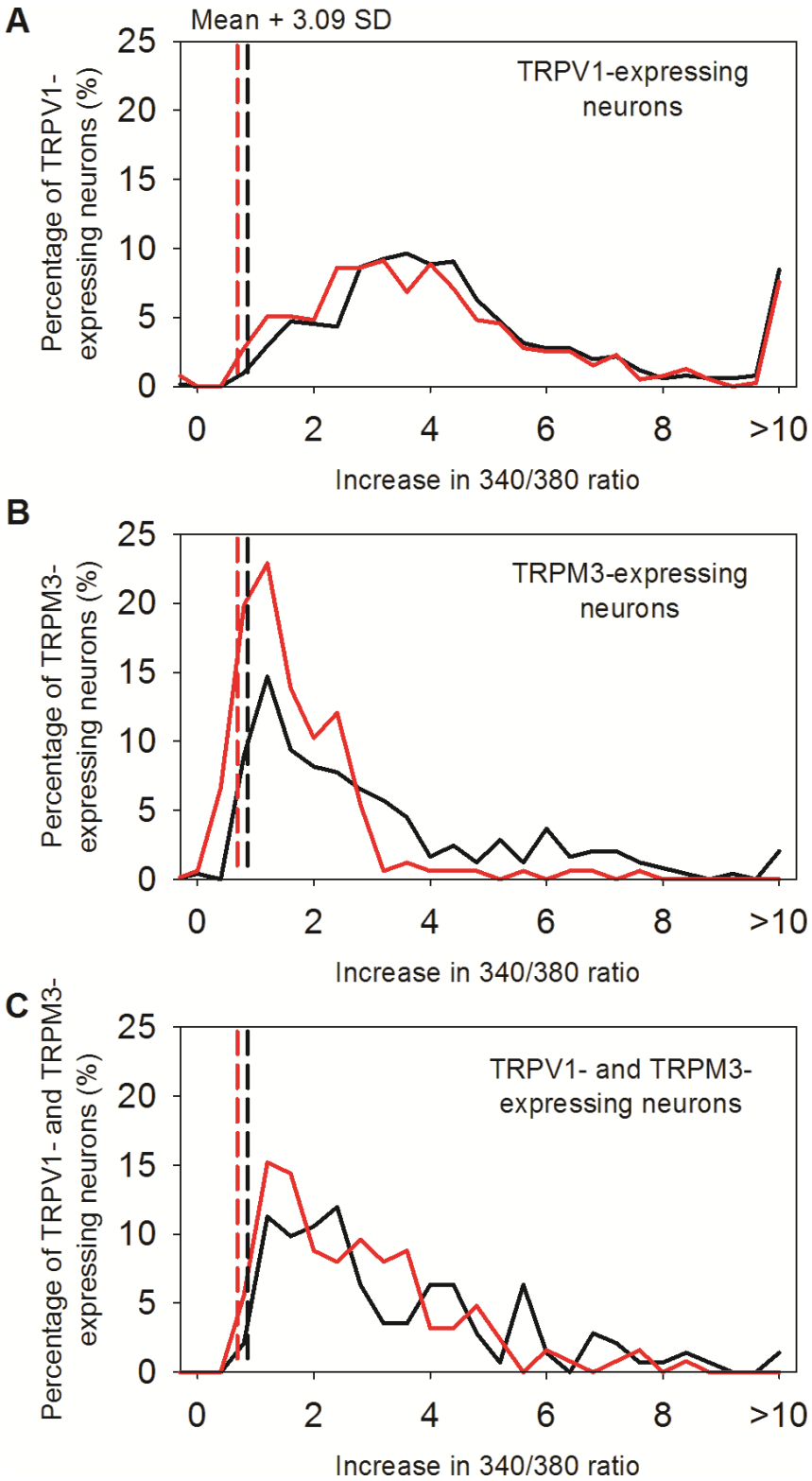
Extended Data Figure 7 | The heat-induced membrane current in autonomic neurons is not gated by membrane voltage. Current–voltage difference relations of a PPG neuron with a voltage ramp starting from a negative potential (inset) show a similar linear heat-induced current to that obtained with reverse voltage ramp (see Fig. 2c). Trace obtained by subtracting current–voltage relations at 36 °C from that at 47 °C. Similar results were obtained in 3 cells on 3 coverslips.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



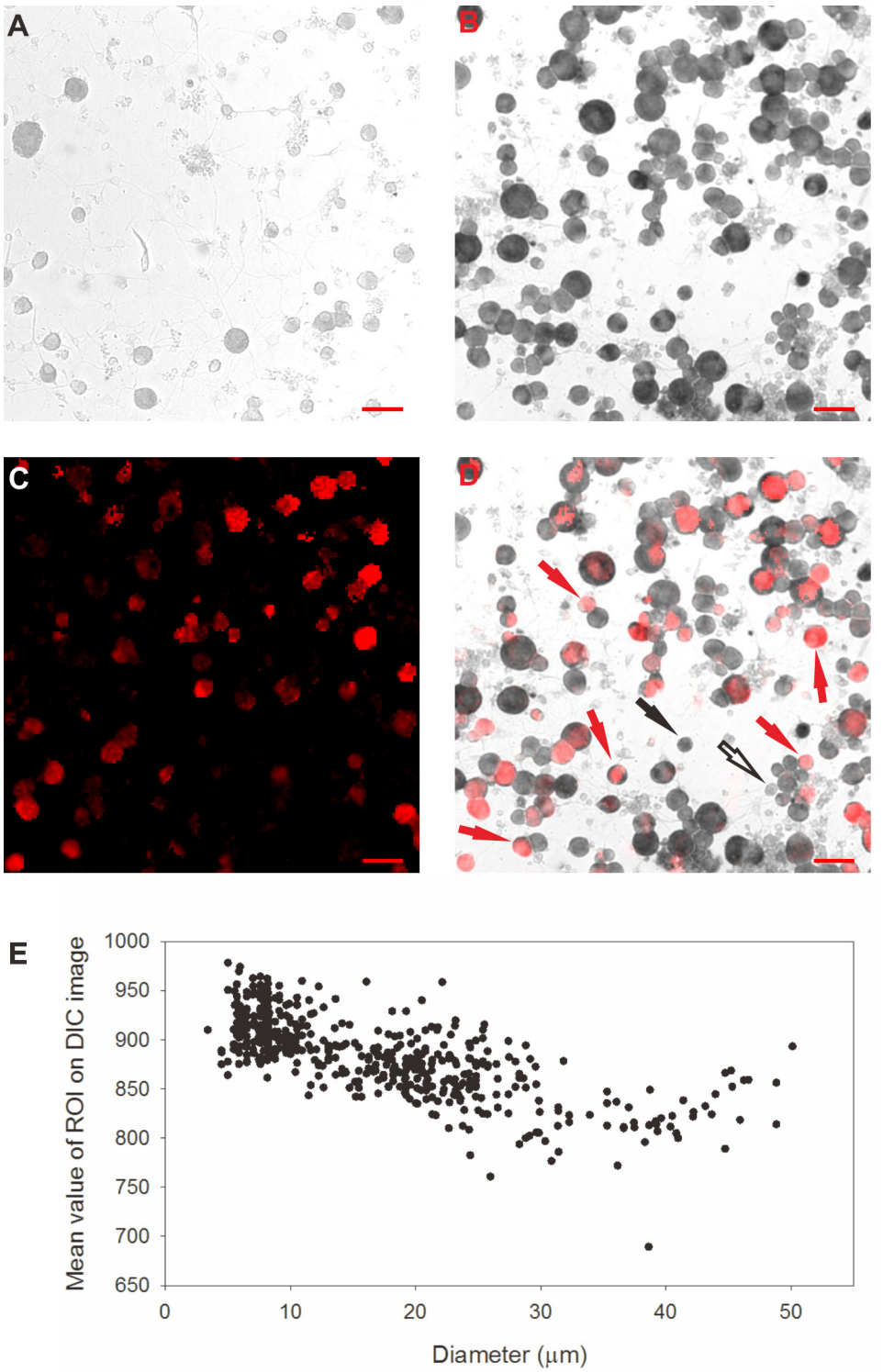
Extended Data Figure 8 | Responses to heat in adrenal-derived MAH and PC12 cell lines, and effects of factors causing differentiation to neuronal phenotype. a, MAH cells. Black indicates maximum increase in $F_{340/380}$ ratio in response to heat (47 °C, $n = 254$) when cultured in dexamethasone (5 μM). No cell responded to TRP agonists, but 27% of cells responded to heat with increase above mean criterion level obtained from glial cells in neuronal cultures (see Fig. 1b). Red indicates similar histogram after 12 days of culture in growth factors (bFGF, CNTF and NGF, see Methods, $n = 170$). No cell responded to TRP agonists; 9% responded to heat. The proportion of heat-sensitive cells was significantly reduced from 27% (69/254) in dexamethasone to 9% (16/170) in growth factors ($P \leq 0.0001$; Fisher's exact test). The 66% reduction in the proportion of heat-sensitive cells was not significantly different from the reduction in *Trpm2* expression caused by differentiation of MAH cells (Table 1; $P = 0.056$; two-tailed unpaired *t*-test). The mean increases in $F_{340/380}$ ratio above the cut-off values (dashed lines) in response to heat were 1.755 ± 0.1255 in dexamethasone ($n = 69$) and 1.420 ± 0.1474 in presence of growth factors ($n = 16$) ($P = 0.2203$; two-tailed unpaired *t*-test). **b**, PC12 cells. Black indicates culture in growth medium (10% horse serum plus 5% fetal bovine serum, $n = 200$). 93% of cells responded to heat with increase above mean criterion level obtained from glial cells in neuronal cultures (see Fig. 1b). Red indicates effect on heat responses of 12 days of culture in NGF (1% horse serum plus 100 ng ml^{-1} NGF, $n = 108$). The proportion of heat-sensitive cells was significantly reduced from 93% (186/200) in growth medium to 46% (50/108) in NGF ($P \leq 0.0001$; Fisher's exact test). We note that a significantly lower expression of mRNA for TRPM2 in differentiated PC12 cells has been reported[31]. The mean increase in $F_{340/380}$ ratio above the cut-off values (dashed lines) in response to heat was significantly reduced from 3.753 ± 0.2431 in growth medium ($n = 186$) to 2.603 ± 0.3104 in NGF ($n = 50$) ($P = 0.0213$; two-tailed unpaired *t*-test). **c**, Temperature thresholds of PC12 cells cultured in growth medium. Top left, temperature protocol. Bottom left, temperature responses of three representative cells. Right, temperature thresholds calculated as in Fig. 1d. Cell numbers and replicates for **a** and **b** were 2 coverslips for each condition imaged. Cell numbers are given above. Replicates of 4 coverslips (MAH cells) and 3 coverslips (PC12 cells) for each condition gave similar results. Cell numbers and replicates for **c** were 165 PC12 cells on one coverslip imaged.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 9 | Effect of deletion of TRPM2 on maximal calcium responses to heat in neurons expressing TRPV1 or TRPM3. **a**, Maximum increases in $F_{340/380}$ ratio in response to a heat ramp from 34 °C to 46 °C in neurons responding only to capsaicin (TRPV1-expressing) from wild-type (black) and *Trpm2*$^{-/-}$ (red) mice. The increase in $F_{340/380}$ ratio in response to heat (above the increase caused by the effect of temperature on fura-2, vertical dotted lines, for method of calculation see Fig. 1b) is not significantly different between wild type and *Trpm2*$^{-/-}$ ($P = 0.1168$, two-tailed Mann–Whitney *U*-test). Details as in Fig. 1. **b**, Neurons responding only to pregnenolone sulphate (PS, TRPM3-expressing) from the same experiments as in **a**. The increases in $F_{340/380}$ ratio in response to heat are significantly reduced by deletion of *Trpm2* (from 6.389 ± 1.225 to 4.411 ± 1.582, $P < 0.0001$, two-tailed Mann–Whitney *U*-test). **c**, Neurons responding to both capsaicin and PS (TRPV1- and TRPM3-expressing). The increases in $F_{340/380}$ ratio in response to heat are not significantly different between wild type and *Trpm2*$^{-/-}$ ($P = 0.0633$; two-tailed Mann–Whitney *U*-test). Cell numbers and replicates for **a**–**c** were 1324 DRG neurons from one wild-type mouse on 4 coverslips and 981 DRG from one *Trpm2*$^{-/-}$ mouse on 4 coverslips imaged. Similar results obtained from 42 additional coverslips from 6 wild-type mice and 8 additional coverslips from one *Trpm2*$^{-/-}$ mouse.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 10 | See next page for caption.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Figure 10 | Correlation between novel heat sensitivity and expression of mRNA for TRPM2. a, Representative DIC transmission image of DRG neurons following *in situ* hybridization with the sense probe as negative control. Non-specific density was linearly dependent on cell diameter (see **e**). Mean + 3.09 s.d. (cumulative probability value of 99.9%) of density as a function of diameter with sense probe (5 μm bins) was used as threshold criterion for significant expression of *Trpm2* in images of antisense hybridization. A similar analysis of non-specific density was carried out for glial cells. **b**, Representative DIC transmission image with antisense probe against *Trpm2*. Using the threshold criterion as function of diameter obtained from sense probe images (see **a**), 89% (1,121/1,250) of DRG neurons but 3% (4/120) of glial cells were positive for TRPM2 mRNA. **c**, Novel heat-sensitive DRG neurons determined using calcium imaging. Increases in $[Ca^{2+}]_i$ ($F_{340/380}$ ratio image, intensity-coded in red) in response to a heat ramp to 46 °C with TRPV1 blocked with AMG9810 (5 μM) and TRPM3 blocked with naringenin (10 μM). **d**, Superimposed image of novel heat-sensitive neurons (red) and *in situ* hybridization using antisense probe. Solid red arrows indicate novel heat-sensitive neurons also positive for TRPM2; solid black arrow shows neuron not responding to heat but positive for TRPM2; open black arrow shows cell negative for TRPM2 (that is, with density below the criterion level obtained from **e**). 42% (92/218) of DRG neurons positive for TRPM2 exhibited novel heat-sensitivity. However only 13% (2/16) of DRG neurons negative for TRPM2 from *in situ* hybridization exhibited novel heat-sensitivity. The percentage of novel heat-sensitive DRG neurons is significantly reduced in TRPM2 negative DRG neurons ($P = 0.0188$; Fisher's exact test). Scale bars, 50 μm. **e**, Density of non-specific label in neurons obtained from hybridization with sense probe (see **a**) depends on cell size. Data used to calculate significance thresholds for neurons in **b**. Cell numbers and replicates for **a** was one coverslip exposed to sense probe used as negative control. Cell numbers and replicates for **b** were all neurons on 5 coverslips measured and one coverslip was measured for TRPM2-positive glial cells. Cell numbers and replicates for **c** and **d** was one coverslip was analysed as in **a** and **b** for the combined calcium imaging and *in situ* hybridization protocol. Cell numbers and replicates for **e** were 500 DRG neurons on one coverslip exposed to the sense probe used to determine the background threshold as a function of cell diameter. Similar *in situ* hybridization results were obtained on 16 additional coverslips.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

LETTER

doi:10.1038/nature19312

The prion protein is an agonistic ligand of the G protein–coupled receptor Adgrg6

Alexander Küffer[1]*, Asvin K. K. Lakkaraju[1]*, Amit Mogha[2], Sarah C. Petersen[2], Kristina Airich[1], Cédric Doucerain[1], Rajlakshmi Marpakwar[1], Pamela Bakirci[1], Assunta Senatore[1], Arnaud Monnard[1], Carmen Schiavi[1], Mario Nuvolone[1], Bianka Grosshans[3], Simone Hornemann[1], Frederic Bassilana[3], Kelly R. Monk[2] & Adriano Aguzzi[1]

Ablation of the cellular prion protein PrP^C leads to a chronic demyelinating polyneuropathy affecting Schwann cells. Neuron-restricted expression of PrP^C prevents the disease[1], suggesting that PrP^C acts *in trans* through an unidentified Schwann cell receptor. Here we show that the cAMP concentration in sciatic nerves from PrP^C-deficient mice is reduced, suggesting that PrP^C acts via a G protein-coupled receptor (GPCR). The amino-terminal flexible tail (residues 23–120) of PrP^C triggered a concentration-dependent increase in cAMP in primary Schwann cells, in the Schwann cell line SW10, and in HEK293T cells overexpressing the GPCR Adgrg6 (also known as Gpr126). By contrast, naive HEK293T cells and HEK293T cells expressing several other GPCRs did not react to the flexible tail, and ablation of Gpr126 from SW10 cells abolished the flexible tail-induced cAMP response. The flexible tail contains a polycationic cluster (KKRPKPG) similar to the GPRGKPG motif of the Gpr126 agonist type-IV collagen[2]. A KKRPKPG-containing PrP^C-derived peptide ($FT_{23–50}$) sufficed to induce a Gpr126-dependent cAMP response in cells and mice, and improved myelination in hypomorphic *gpr126* mutant zebrafish (*Danio rerio*). Substitution of the cationic residues with alanines abolished the biological activity of both $FT_{23–50}$ and the equivalent type-IV collagen peptide. We conclude that PrP^C promotes myelin homeostasis through flexible tail-mediated Gpr126 agonism. As well as clarifying the physiological role of PrP^C, these observations are relevant to the pathogenesis of demyelinating polyneuropathies—common debilitating diseases for which there are limited therapeutic options.

Neuronal ablation of *Prnp* triggers chronic demyelinating polyneuropathy (CDP)[1], suggesting that Schwann cells bear a PrP^C receptor. We therefore assessed the binding of full-length PrP^C (recPrP, residues 23–231), the flexible tail (FT; residues 23–110), or its refolded globular domain (GD; residues 121–231), to primary Schwann cell (PSC) cultures from $Prnp^{ZH1/ZH1}$ sciatic nerves[3] using POM1 and POM2 antibodies against epitopes of PrP^C (ref. 4). Both recPrP and FT, but not GD, were found to stain PSCs (Extended Data Fig. 1a).

Using transcription activator-like effector nucleases, we generated a *Prnp*-ablated subclone (termed $SW10_{\Delta PrP}$) of the SW10 Schwann cell line (Extended Data Fig. 1b, c). Again, recPrP and FT, but not GD, adhered to $SW10_{\Delta PrP}$ cells (Fig. 1a and Extended Data Fig. 1d, e). Neither recPrP nor FT adhered to the $Prnp^{-/-}$ hippocampal cell line HpL[5] (Extended Data Fig. 1f), suggesting that binding was specific to Schwann cells. We next measured the binding of synthetic PrP-derived peptides (2 μM, 20 min) to $SW10_{\Delta PrP}$ cells (Fig. 1b). $FT_{23–50}$, but neither $FT_{39–66}$ nor any of the C-proximal peptides (Extended Data Table 1), showed binding to $SW10_{\Delta PrP}$ cells (Fig. 1c), suggesting that residues 23–38 are essential for this interaction.

We treated SW10 cells with trypsin (2.5% w/v, 10 min) to degrade membrane-resident proteins. Following trypsin inactivation, we added non-trypsinized SW10 cells labelled with the cell-tracking dye Deep Red. Haemagglutinin (HA)-tagged $FT_{23–50}$ (2 μM) was added to the cells, and binding was monitored with anti-HA antibodies. $FT_{23–50}$ bound to 51% of the non-trypsinized cells and 5% of the trypsinized cells (Extended Data Fig. 1g), suggesting that there is a surface FT receptor on Schwann cells. Cytofluorimetry revealed no altered binding of HA-tagged $FT_{23–50}$ to SW10 cells treated with phosphoinositol phospholipase C (PI-PLC, 30 min at 37 °C), indicating that binding did not require glycophosphatidylinositol (GPI)-anchored surface proteins, whereas binding of POM2 was greatly reduced, indicating that PrP^C had been stripped from cell surfaces (Extended Data Fig. 1h).

Might the FT alter cAMP signalling in Schwann cells? Sciatic nerves taken from $Prnp^{ZH1/ZH1}$ and wild-type mice at 4 days of age showed similar cAMP levels (Extended Data Fig. 2a). Sciatic nerves

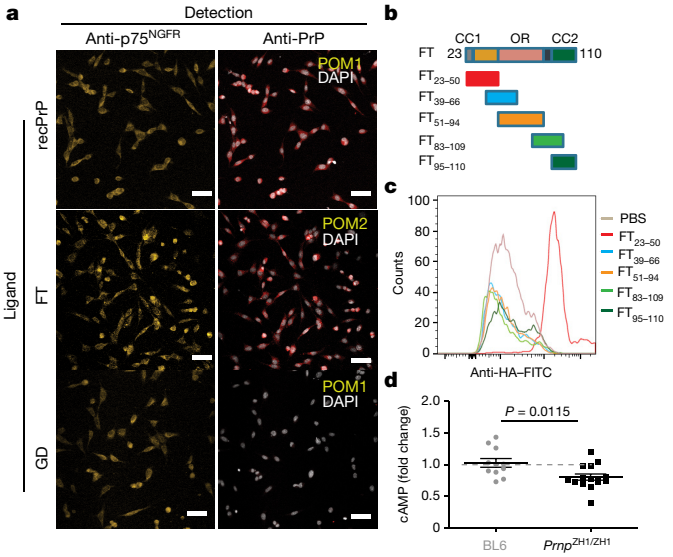


Figure 1 | Schwann cells selectively bind the FT of PrP^C. a, *Prnp*-ablated SW10 cells ($SW10_{\Delta PrP}$) were exposed to recombinant PrP^C (recPrP), FT, or GD (ligand). PrP^C and FT, but not GD, adhered to Schwann cells (red: POM1 and POM2 antibodies; grey: DAPI). $p75^{NGFR}$ antibodies identified Schwann cells. Scale bars, 26 μm. **b**, FT-derived peptides and PrP^C domains. CC1 and CC2, charge clusters 1 and 2; OR, octapeptide repeats. Peptides are colour-coded as in **c**. **c**, $SW10_{\Delta PrP}$ cells were exposed to FT-derived peptides (2 μM, 20 min) carrying a C-terminal HA tag. Flow cytometry showed strong binding by peptide $FT_{23–50}$. **d**, Sciatic nerves from $Prnp^{ZH1/ZH1}$ mice displayed lower cAMP than those from wild-type BL6 mice. Dots represent individual mice (11–15 mice per group). Error bars show s.e.m. Unpaired Student's *t*-test was used for statistical analysis. Data (**a**, **c**) are representative of three biological replicates.

[1]Institute of Neuropathology, University of Zurich, CH-8091 Zürich, Switzerland. [2]Washington University School of Medicine, Department of Developmental Biology and Hope Center for Neurological Disorders, 660 South Euclid Avenue, Campus Box 8103, St. Louis, Missouri 63110, USA. [3]Novartis Institutes for Biomedical Research, CH-4056 Basel, Switzerland.
*These authors contributed equally to this work.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

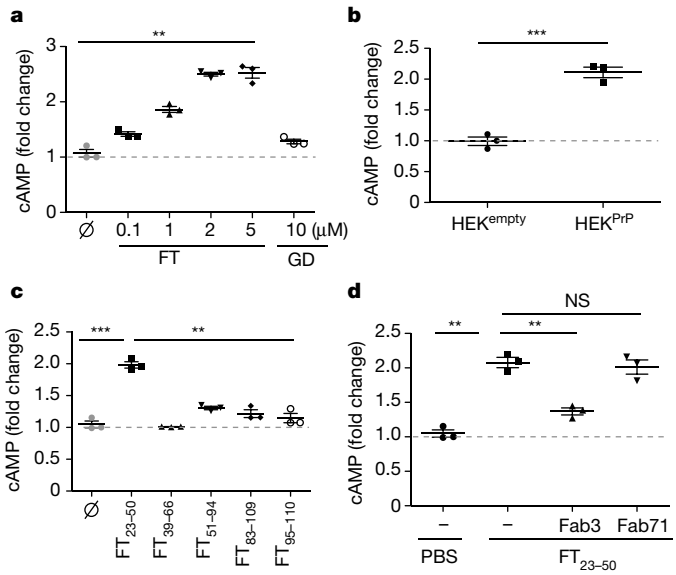


Figure 2 | The FT fragment elicits a concentration-dependent cAMP response. a, Primary $Prnp^{ZH1/ZH1}$ Schwann cells were treated (20 min) with increasing concentrations of recombinant FT or with 10 μM GD. Ø, untreated cells. cAMP levels were determined in cell lysates (5×10^5 cells per assay). Addition of FT, but not of GD, induced a concentration-dependent cAMP response in Schwann cells. NS, not significant; $^{**}P < 0.01$; $^{***}P < 0.001$. **b**, cAMP concentrations in primary $Prnp^{ZH1/ZH1}$ Schwann cell cultures exposed to medium conditioned by HEK293T cells overexpressing wild-type murine PrP^C (HEK^{PrP}, right) or a non-coding vector (HEK^{empty}, left) as control. FT-containing medium resulted in cAMP induction. **c**, Synthetic peptides (27–44 residues) were added to $SW10_{\Delta PrP}$ cells (2 μM, 20 min). Only $FT_{23–50}$ induced cAMP. **d**, $FT_{23–50}$ was preincubated with a twofold molar excess of miniantibodies Fab3 or Fab71, and added to $SW10_{\Delta PrP}$ cells (20 min). Preincubation with Fab3, but not with Fab71, significantly quenched the FT-dependent cAMP spike. Error bars show s.e.m. Panels depict independent triplicates; unpaired Student's *t*-test was used for statistical analysis.

taken from 4-week-old $Prnp^{ZH1/ZH1}$ mice showed a trend towards decreased cAMP levels (Extended Data Fig. 2b). When taken from $Prnp^{ZH1/ZH1}$ mice at 12–16 weeks, when CDP is incipient[1], sciatic nerves exhibited significantly lower cAMP levels than those taken from wild-type mice (Fig. 1d; $P = 0.0115$). Sciatic nerve lysates from strictly isogenic C57BL/6J $Prnp^{ZH3/ZH3}$ mice[6] (10–16 weeks old) also displayed significantly lower levels of cAMP than did those from wild-type mice (Extended Data Fig. 2c).

Sciatic nerves from 12–16-week-old *tg*NSE-PrP mice, which express PrP^C in neurons, showed cAMP levels similar to those of wild-type mice, whereas sciatic nerves from *tg*MBP-PrP mice, which express PrP^C in Schwann cells and suffer from CDP[1], showed lower cAMP levels (Extended Data Fig. 2d). Moreover, when treated with FT (0.1–5 μM, 20 min), $Prnp^{ZH1/ZH1}$ PSC and SW10 cells displayed a concentration-dependent increase in cAMP (Fig. 2a and Extended Data Fig. 2e) with an EC_{50} of 860 nM (Extended Data Fig. 2f).

We transfected HEK293T cells, which express little endogenous PrP^C, with a plasmid expressing murine PrP^C or with an empty plasmid (HEK^{PrP} or HEK^{empty}, respectively). Immunoprecipitation showed that HEK^{PrP} cells released soluble FT (approximately 37 ng FT per ml; Extended Data Fig. 3a, b). Exposure to conditioned medium from HEK^{PrP} cells, but not from HEK^{empty} cells, raised cAMP levels in PSC cultures generated from $Prnp^{ZH1/ZH1}$ mice (Fig. 2b). Spent medium from PSC cultures and sciatic nerves from 10-week-old C57BL/6 mice did not contain any FT, suggesting that little FT is released by Schwann cells (Extended Data Fig. 3c, d).

To find out whether the FT contains motifs responsible for inducing the cAMP response, we treated $SW10_{\Delta PrP}$ cells with the same synthetic peptides used previously to assess binding to Schwann cells. Peptide $FT_{23–50}$ induced a cAMP response, whereas $FT_{34–69}$ and all C-proximal peptides were inactive (Fig. 2c). These results suggest that residues 23–33, which contain the lysine-rich charge cluster 1 (CC1), represent the biologically active region of the FT. To test this prediction, we incubated $FT_{23–50}$ with monovalent recombinant phage-derived miniantibodies recognizing the CC1 (Fab3) or octapeptide repeats (Fab71) of PrP^C. Preincubation with Fab3, but not with Fab71, significantly quenched the ability of $FT_{23–50}$ to induce an increase in cAMP in $SW10_{\Delta PrP}$ cells (Fig. 2d).

Peripheral nervous system myelination is controlled by Gpr126, an adhesion GPCR expressed by Schwann cells[7,8]. We therefore generated HEK293T cells stably overexpressing human Gpr126 or, as controls, Gpr124 or Gpr176 (denoted HEK^{Gpr126}, HEK^{Gpr124}, and HEK^{Gpr176}, respectively) with a C-terminal V5 epitope tag[9] (Extended Data Fig. 3e). Flow cytometry revealed that $FT_{23–50}$ bound to HEK^{Gpr126} cells but not to HEK^{Gpr124} or HEK^{Gpr176} cells (Extended Data Fig. 3f, g). Moreover, exposure to $FT_{23–50}$ (20 min, ≥500 nM) increased cAMP in HEK^{Gpr126} cells but not in HEK^{Gpr124} or HEK^{Gpr176} cells (Fig. 3a). We then treated HEK293T, HEK^{Gpr124}, and HEK^{Gpr126} cells with recombinant FT or GD (20 min, 2 μM). FT was detected only in lysates from HEK^{Gpr126} cells immunoprecipitated with anti-V5 antibodies, and GD was not detected in lysates from any of the cells (Extended Data Fig. 4a).

Next, we generated clonal SW10 cells devoid of endogenous Gpr126 (designated $SW10_{\Delta Gpr126}$). When treated with $FT_{23–50}$, SW10 (but not $SW10_{\Delta Gpr126}$) cells reacted with increased cAMP levels. A modified $FT^{K \to A}_{23–50}$ peptide (with lysine residues replaced with alanines) was ineffective in binding cells and inducing cAMP (Fig. 3b). We then treated $SW10_{\Delta Gpr126}$ cells transfected with human Gpr126, Gpr124, Gpr176 or Gpr56 with $FT_{23–50}$. Only Gpr126-transfected cells showed a cAMP response (Extended Data Fig. 4b) similar to that of naive SW10

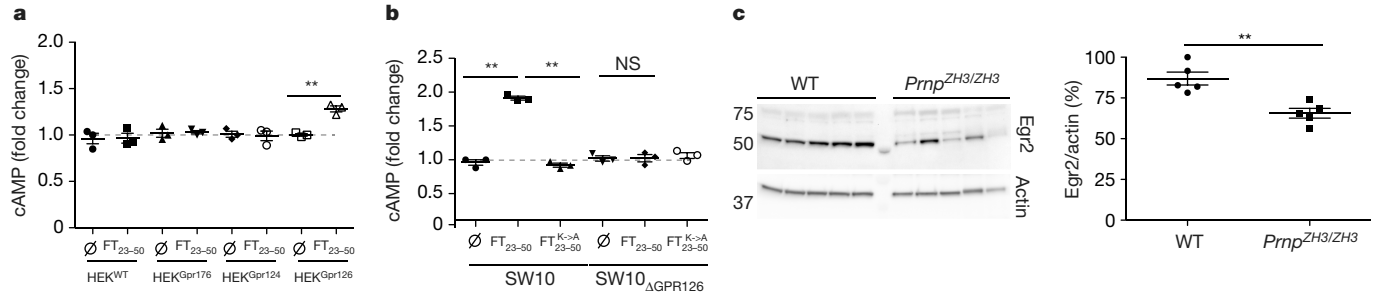


Figure 3 | FT-dependent cAMP signalling in Gpr126-ablated Schwann cells. a, Intracellular cAMP in wild-type HEK293T cells and in Gpr176-, Gpr124- and Gpr126-overexpressing cells exposed to $FT_{23–50}$ (0.5 μM, 20 min). Only HEK^{Gpr126} cells showed an increase in cAMP. **b**, Wild-type (left) and Gpr126-ablated (right) SW10 cells were exposed to $FT_{23–50}$ (2 μM, 20 min). SW10 cells, but not $SW10_{\Delta Gpr126}$ cells, responded to $FT_{23–50}$ with a cAMP spike. Moreover, SW10 cells did not respond to alanine-substituted $FT_{23–50}$ ($FT^{K \to A}_{23–50}$). **c**, Protein was isolated from wild-type or $Prnp^{ZH3/ZH3}$ sciatic nerves (13-week-old female mice) and western blots were probed for Egr2 and actin. Densitometry (below) showed reduced Egr2 in $Prnp^{ZH3/ZH3}$ nerves ($P = 0.028$). For uncropped gels see Supplementary Information File 1. Error bars show s.e.m. Panels depict independent triplicates; unpaired Student's *t*-test was used for analysis. $^{**}P < 0.01$.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

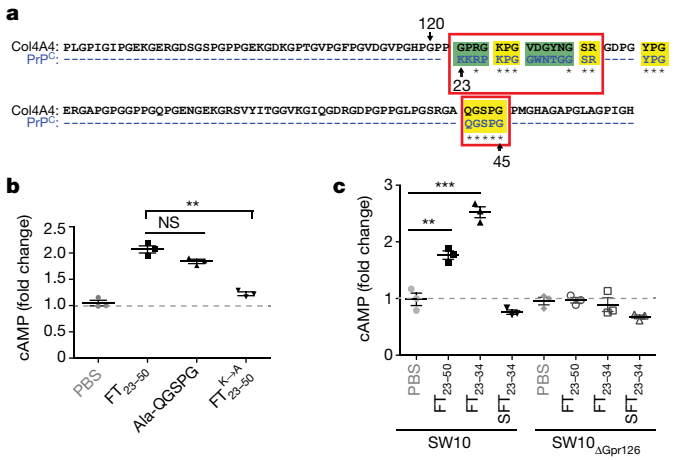


Figure 4 | FT and collagen-IV share a cAMP-inducing domain.
a, Sequence alignment revealed two regions of similarity between the FT and Col4 (red boxes). Yellow and green shades represent high and moderate similarity, respectively. Dotted line, non-homologous residues; asterisks, identical residues. **b**, SW10$_{\Delta PrP}$ cells were treated with synthetic FT_{23-50} or modified version of FT_{23-50} in which the KKRPK or QGSPG motifs were replaced with alanines (2 μM, 20 min). Alanine substitution of KKRPK (peptide $FT^{K\rightarrow A}_{23-50}$), but not of QGSPG, abrogated cAMP induction. **c**, SW10$_{\Delta PrP}$ and SW10$_{\Delta Gpr126}$ cells were exposed (2 μM, 20 min) to the synthetic peptides FT_{23-50}, FT_{23-34} or SFT_{23-34} (peptide containing scrambled amino acid sequence of FT_{23-34}). FT_{23-50} and FT_{23-34} induced cAMP in SW10$_{\Delta PrP}$ cells but not SW10$_{\Delta Gpr126}$ cells. Error bars show s.e.m. Panels depict independent triplicates; unpaired Student's *t*-test was used for analysis. $^{**}P < 0.01$; $^{***}P < 0.001$.

cells, indicating that the tag did not affect the function of Gpr126 (Extended Data Fig. 4c). When treated with conditioned medium from HEK^{PrP} or HEK^{empty} cells, SW10 but not SW10$_{\Delta Gpr126}$ cells responded with a cAMP spike (Extended Data Fig. 4d). Moreover, FT adsorption was reduced in SW10$_{\Delta Gpr126}$ cells (Extended Data Fig. 4e).

We then administered FT_{23-50} (20 min, 2 μM) to HEK^{Gpr126} cells and HEK293(H) cells (a clonal variant of HEK293T cells with superior growth and transfection efficiencies) transfected with plasmids encoding human Gpr56, Gpr64, Gpr133, or Gpr97. Only Gpr126-expressing cells showed a cAMP response (Extended Data Fig. 4f). The magnitude of the cAMP response was not enhanced by increasing the amount of transfected plasmid, suggesting that other signalling components became limiting (Extended Data Fig. 5a). There was no cAMP induction in $Prnp^{ZH1/ZH1}$ cerebellar granule neuronal cultures treated with FT_{23-50} or $FT^{K\rightarrow A}_{23-50}$ (Extended Data Fig. 5b), as expected from the minimal Gpr126 expression in the brain[10]. The FT is released from PrP^C by metalloproteases[11]; after treatment with the metalloprotease inhibitor TAPI-2, HEK^{PrP}-conditioned medium contained significantly less FT than untreated medium (Extended Data Fig. 5c, d) and displayed reduced cAMP-inducing activity (Extended Data Fig. 5c).

Egr2 (also known as Krox-20) controls the expression of myelin genes and has been implicated in myelin maintenance[12]. Egr2 expression was decreased in sciatic nerves from 13-week-old $Prnp^{ZH3/ZH3}$ mice ($P < 0.05$; Fig. 3c), and recombinant FT (2 μM, 24 h) activated Egr2-dependent luciferase expression in SW10 cells (Extended Data Fig. 5e). Similarly, *Egr2* transcription was upregulated in primary Schwann cells treated with recombinant FT (2 μM, 1 h) (Extended Data Fig. 5f). Also, Akt phosphorylation increased 5 min after treatment with recombinant FT (2 μM) and peaked at 10 min in SW10$_{\Delta PrP}$ but not SW10$_{\Delta Gpr126}$ cells (Extended Data Fig. 5g). The integrity of SW10 cells and their subclones was confirmed by the expression of myelin genes (Extended Data Fig. 6a).

We identified two regions of similarity between FT (KKRPKPG and QGSPG) and the Gpr126 ligand, type-IV collagen (Col4)[2] (GPRGKPG and QGSPG; Fig. 4a). Replacement of the conserved cationic residues with alanines (KKRPKPG → AAAPAPG), but not other substitutions, abrogated cAMP induction in SW10$_{\Delta PrP}$ cells (Fig. 4b); treatment with FT_{23-34} (2 μM, 20 min), which contains KKRPKPG, sufficed to induce cAMP in SW10$_{\Delta PrP}$ but not in SW10$_{\Delta Gpr126}$ cells (Fig. 4c). We next generated murine PrP^C mutants containing alanine substitutions in either of the two conserved motifs. After transient transfection, both mutants were highly expressed by HEK293T cells (Extended Data Fig. 6b), and cleaved FT was recovered in the medium (Extended Data Fig. 6c). When applied to SW10$_{\Delta PrP}$ cells, conditioned medium from HEK293T cells expressing wild-type or QGSPG-mutated PrP^C ($HEK^{PrP}_{\sum QGSPG}$) induced cAMP, whereas medium from cells expressing KKRPK-mutated PrP^C ($HEK^{PrP}_{\sum KKRPK}$) did not (Extended Data Fig. 6d). We then generated 21-mer peptides bearing the corresponding Col4 sequence (GPRGKPG) or an alanine-substituted variant (AAAGAAG). The native Col4 peptide (8 μM), but not the mutated peptide, induced cAMP in SW10$_{\Delta PrP}$ cells (Extended Data Fig. 6e).

In zebrafish, Gpr126 controls myelination; however, $Prnp^{-/-}$ mice have no obvious myelination defects but develop a late-onset peripheral neuropathy with demyelination, onion bulb formation (Fig. 5a and Extended Data Fig. 7a), decreased Remak bundles containing 10 or more axons and increased bundles containing fewer than 10 axons (Extended Data Fig. 7b), indicative of impaired axon–Schwann cell interactions. To test whether Gpr126 dysfunction could cause late-onset phenotypes in mice, we examined sciatic nerves from Dhh^{Cre}::$Gpr126^{fl/fl}$ mice, in which *Gpr126* is specifically deleted in Schwann cells from approximately embryonic day (E)12.5 onwards[13,14]. At one year of age, these mice showed neuropathic traits similar to those of $Prnp^{-/-}$ mice including reduced bundles containing >20 unmyelinated axons, increased bundles containing <10 axons, abnormal cytoplasmic Schwann cell protrusions[1] and onion bulbs (Fig. 5b and Extended Data Fig. 7c, d).

The $gpr126^{st63}$ zebrafish mutant has a point mutation that reduces Gpr126 signalling and shows decreased myelin basic protein (Mbp) expression by Schwann cells of the posterior lateral line nerve (PLLn) (Fig. 5c and Extended Data Fig. 8a), which can be rescued by Gpr126 activators[8,15,16]. When applied to $gpr126^{st63}$ zebrafish larvae at 50–55 h post-fertilization (hpf), FT_{23-50} (20 μM) increased Mbp expression in the PLLn at 5 days post-fertilization (dpf) compared to DMSO-treated larvae (Fig. 5c, d; $P < 0.05$). We also treated $gpr126^{st49}$ larvae, which encode a truncated Gpr126 incapable of G_s signalling[8,16]. Mbp was barely detectable in both FT_{23-50}- and DMSO-treated mutant $gpr126^{st49}$ larvae (Fig. 5c and Extended Data Fig. 8b).

We next administered intravenous FT_{23-50} or $FT^{K\rightarrow A}_{23-50}$ peptide (600 μg) to 10–16-week-old $Prnp^{ZH3/ZH3}$ and wild-type mice ($n = 8$ per group, all littermates). FT_{23-50}, but not $FT^{K\rightarrow A}_{23-50}$, induced cAMP elevation in both $Prnp^{ZH3/ZH3}$ and $Prnp^{ZH1/ZH1}$ sciatic nerves 20 min after the injection; cAMP reached levels similar to those of BL6 mice injected with $FT^{K\rightarrow A}_{23-50}$ (Fig. 5e and Extended Data Fig. 8c). FT_{23-50} also induced a cAMP spike in the heart, which expresses Gpr126 (Fig. 5f and Extended Data Fig. 8d) but not in the kidney or brain (Extended Data Fig. 8e, f), which do not express Gpr126 (ref. 17). Finally, FT_{23-50} but not $FT^{K\rightarrow A}_{23-50}$ (600 μg) induced a robust cAMP spike in sciatic nerves from $Gpr126^{fl/fl}$ but not Dhh^{Cre}::$Gpr126^{fl/fl}$ mice (Fig. 5g).

Here we report, to our knowledge, the first molecular elucidation of a phenotype caused by PrP^C deficiency. While Gpr126 is crucial for peripheral nerve myelination during development, the late-onset phenotype of *Prnp*-ablated mice suggests that it has additional roles in myelin maintenance. $Gpr126^{-/-}$ mice exhibit drastic hypomyelination, but the $Prnp^{-/-}$ phenotype is relatively mild and late-onset, akin to peripheral neuropathies, whose prevalence is 2.4% generally and 8% among the aged[18]. Perhaps Gpr126 provides basal signalling in $Prnp^{-/-}$ mice through its interaction with type-IV collagen and/or laminin-211 (refs 2, 16), which may not suffice for long-term myelin maintenance. The Gpr126-agonistic properties of systemically administered FT

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

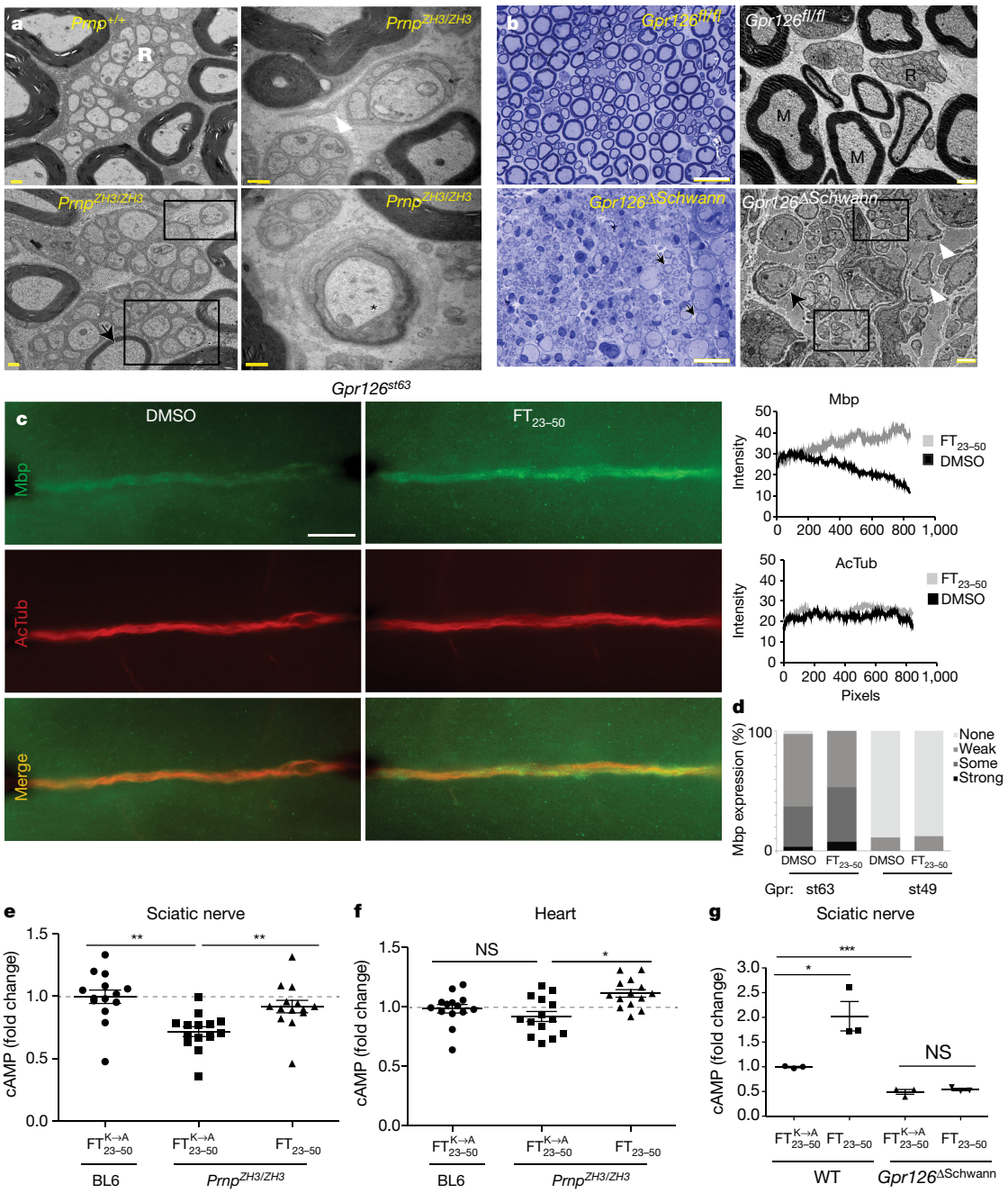


Figure 5 | Myelinotrophic effect of FT in zebrafish and mice. **a**, Transmission electron micrographs of sciatic nerves from 14-month-old wild-type BL6 and $Prnp^{ZH3/ZH3}$ mice ($n = 3$–4). Thinly myelinated axons (black arrow), loss of axon–Schwann cell interaction (boxes), abnormal cytoplasmic Schwann cell protrusions (white arrowhead) and initial onion bulb formation (asterisk) were observed in $Prnp^{ZH3/ZH3}$ mice. Scale bars, 500 nm. **b**, Neuropathic phenotype of nerves from one-year-old Dhh^{Cre}::$Gpr126^{fl/fl}$ mice. Left, toluidine blue-stained sections of sciatic nerves from control $Gpr126^{fl/fl}$ (phenotypically wild-type) and Dhh^{Cre}::$Gpr126^{fl/fl}$ ($Gpr126^{\Delta Schwann}$) mice. $Gpr126^{fl/fl}$ nerves were well-myelinated ($n = 3/3$ mice), whereas $Gpr126^{\Delta Schwann}$ nerves exhibited myelin loss with readily apparent onion bulb-like structures (arrows; $n = 3/3$ mice). Right panels: Transmission electron micrographs of sciatic nerves from $Gpr126^{fl/fl}$ and $Gpr126^{\Delta Schwann}$ mice. Myelinated axons (M) and Remak bundles (R) were found in $Gpr126^{fl/fl}$ mouse sciatic nerves ($n = 3/3$ mice). Numerous defects were observed in $Gpr126^{\Delta Schwann}$ mouse sciatic nerves ($n = 3/3$ mice) including onion bulbs (black arrow), abnormal cytoplasmic protrusions (white arrowheads), and loss of axon–Schwann cell interactions (boxes) similar to those seen in $Prnp^{ZH3/ZH3}$ mice. Scale bars: 20 μm (left), 2 μm (right). **c**, $gpr126^{st63}$ hypomorphic mutant zebrafish larvae were treated with vehicle (DMSO) or $FT_{23–50}$ (20 μM) at 50–55 hpf and the posterior lateral line nerve was immunostained at 5 dpf for myelin basic protein (Mbp, green). AcTub: acetylated tubulin (red) labelling axons. Scale bar, 20 μm. The intensity of immunofluorescence was assessed by morphometry (right). FT treatment enhanced Mbp immunofluorescence without affecting AcTub. **d**, Mbp expression was scored in larvae treated with $FT_{23–50}$ or vehicle (DMSO). $FT_{23–50}$ treatment resulted in a higher proportion of rescued (some and strong) Mbp expression in $gpr126^{st63}$ (53% versus 34%) but not in $gpr126^{st49}$ larvae ($P < 0.05$, Fisher's two-tailed exact test). NS, not significant. $n \geq 25$ larvae per replicate treatment. **e**, $Prnp^{ZH3/ZH3}$ and BL6 mice were intravenously injected with either $FT_{23–50}$ or its non-charged analogue $FT^{K \to A}_{23–50}$ (600 μg per animal, 20 min). $Prnp^{ZH3/ZH3}$ sciatic nerves showed a significant cAMP increase after injection with $FT_{23–50}$, but not $FT^{K \to A}_{23–50}$. $FT_{23–50}$-treated $Prnp^{ZH3/ZH3}$ sciatic nerves reached cAMP levels similar to those of BL/6 mice injected with $FT^{K \to A}_{23–50}$. Each dot represents an individual animal. **f**, Heart cAMP levels were also increased in $FT_{23–50}$-injected mice. **g**, Control $Gpr126^{fl/fl}$ (WT) and Dhh^{Cre}::$Gpr126^{fl/fl}$ mutant ($Gpr126^{\Delta Schwann}$) mice were intravenously injected with either $FT_{23–50}$ or $FT^{K \to A}_{23–50}$ (600 μg per animal). Sciatic nerves were isolated 20 min after injection. $FT_{23–50}$ elicited a significant cAMP increase in wild-type mice, but not in $Gpr126^{\Delta Schwann}$ mice. $FT^{K \to A}_{23–50}$ injection did not alter cAMP levels ($n = 3$). Error bars show s.e.m. Unpaired Student's *t*-test was used for analysis. $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

peptides suggest that soluble ligands may be beneficial in diseases caused by hypomorphic Gpr126 mutations[19], and perhaps also in other hereditary motor–sensory neuropathies.

A moderately conserved homology region between PrP^C and Col4 proved essential to the biological activity of both proteins, whereas laminin-211 might activate Gpr126 through different mechanisms. Although Gpr126 is not expressed in the central nervous system, certain mutants of PrP^C cause myelin pathology *in vivo*[20]. This observation raises the question of whether inappropriate GPCR activation may play a role in prion diseases within the CNS.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 22 December 2015; accepted 19 July 2016.
Published online 8 August 2016.

1. Bremer, J. *et al.* Axonal prion protein is required for peripheral myelin maintenance. *Nat. Neurosci.* **13,** 310–318 (2010).
2. Paavola, K. J., Sidik, H., Zuchero, J. B., Eckart, M. & Talbot, W. S. Type IV collagen is an activating ligand for the adhesion G protein-coupled receptor GPR126. *Sci. Signal.* **7,** ra76 (2014).
3. Büeler, H. *et al.* Normal development and behaviour of mice lacking the neuronal cell-surface PrP protein. *Nature* **356,** 577–582 (1992).
4. Polymenidou, M. *et al.* The POM monoclonals: a comprehensive set of antibodies to non-overlapping prion protein epitopes. *PLoS One* **3,** e3872 (2008).
5. Kuwahara, C. *et al.* Prions prevent neuronal cell-line death. *Nature* **400,** 225–226 (1999).
6. Nuvolone, M. *et al.* Strictly co-isogenic C57BL/6J-$Prnp^{-/-}$ mice: A rigorous resource for prion science. *J. Exp. Med.* **213,** 313–327 (2016).
7. Pogoda, H. M. *et al.* A genetic screen identifies genes essential for development of myelinated axons in zebrafish. *Dev. Biol.* **298,** 118–131 (2006).
8. Monk, K. R. *et al.* A G protein-coupled receptor is essential for Schwann cells to initiate myelination. *Science* **325,** 1402–1405 (2009).
9. Kasof, G. M. & Gomes, B. C. Livin, a novel inhibitor of apoptosis protein family member. *J. Biol. Chem.* **276,** 3238–3246 (2001).
10. Zhang, Y. *et al.* An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J. Neurosci.* **34,** 11929–11947 (2014).
11. Altmeppen, H. C. *et al.* The sheddase ADAM10 is a potent modulator of prion disease. *eLife* **4,** e04260 (2015).
12. Decker, L. *et al.* Peripheral myelin maintenance is a dynamic process requiring constant Krox20 expression. *J. Neurosci.* **26,** 9771–9779 (2006).
13. Jaegle, M. *et al.* The POU proteins Brn-2 and Oct-6 share important functions in Schwann cell development. *Genes Dev.* **17,** 1380–1391 (2003).
14. Mogha, A. *et al.* Gpr126 functions in Schwann cells to control differentiation and myelination via G-protein activation. *J. Neurosci.* **33,** 17976–17985 (2013).
15. Liebscher, I. *et al.* A tethered agonist within the ectodomain activates the adhesion G protein-coupled receptors GPR126 and GPR133. *Cell Reports* **9,** 2018–2026 (2014).
16. Petersen, S. C. *et al.* The adhesion GPCR GPR126 has distinct, domain-dependent functions in Schwann cell development mediated by interaction with laminin-211. *Neuron* **85,** 755–769 (2015).
17. Uhlén, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347,** 1260419 (2015).
18. Ishida, C., Okino, S., Kitamoto, T. & Yamada, M. Involvement of the peripheral nervous system in human prion diseases including dural graft associated Creutzfeldt-Jakob disease. *J. Neurol. Neurosurg. Psychiatry* **76,** 325–329 (2005).
19. Kou, I. *et al.* Genetic variants in GPR126 are associated with adolescent idiopathic scoliosis. *Nat. Genet.* **45,** 676–679 (2013).
20. Radovanovic, I. *et al.* Truncated prion protein and Doppel are myelinotoxic in the absence of oligodendrocytic PrPC. *J. Neurosci.* **25,** 4879–4888 (2005).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank R. Moos, P. Schwarz, C. Tiberi, C. Sturzenegger, V. Escalante, M. Delic, C. Johnson and Z. Spence for technical help; J. Bremer for discussions; V. Beck, M.-A. Wulf, K. Frontzek, T. Johnson, T. O'Connor, A. Lau and C. Rehwald for help in performing experiments; and X. Piao for antibodies against Gpr126. A.A. is the recipient of support from an Advanced Grant of the European Research Council (Prion2020), a European Union Framework 7 Grant (NEURINOX), the Swiss National Research Foundation, Sinergia grant #147660, the Clinical Research Priority Programs 'Small RNAs' and 'Human Hemato-Lymphatic Diseases', SystemsX.ch, the EU Joint Programme on Neurodegenerative Disease Research (JPND) CureALS and REfrAME, and the Novartis Research Foundation. A.K. was supported by an MD/PhD fellowship from the Swiss National Science Foundation. A.K.K.L. is supported by a grant from the Synapsis Foundation. S.H. is a recipient of a SystemsX.ch grant. This work was supported by NIH grants F32 NS087786 to S.C.P. and NS079445 to K.R.M.

Author Contributions A.A., A.K., K.R.M. and A.K.K.L. designed experiments, analysed data and wrote the manuscript. A.K. generated recombinant proteins, SW10ΔPrP and $SW10_{GPR126}$ cell lines, and performed cAMP, q-PCR, FACS and luciferase assays. A.K.K.L. performed cAMP assays, IP, western blots, immunofluorescence, and FACS. C.D. contributed to immunofluorescence, extracted sciatic nerves from mice, performed intravenous injections and performed repetitions of cAMP and FACS assays. A.M. performed all Schwann cell *Gpr126* mutant mouse experiments and S.C.P. performed all zebrafish experiments, which were analysed by A.M., S.C.P., K.R.M., and A.A. R.M. performed cAMP assays, western blots, and FACS analyses. K.A. contributed to cAMP assays and western blots. P.B. performed the experiments shown in Fig. 1c and repeated the cAMP assay for Fig. 5b–d. A.S. generated Fab3 and Fab71 antibodies. A.M., C.S., and K.A. contributed to Fig. 2b, g, and a, respectively. M.N. generated ZH3 mice and performed electron microscopy. B.G. and F.B. generated GPR-overexpressing HEK cell lines. S.H. supervised generation of recombinant proteins, designed synthetic peptides, and contributed to writing the manuscript. All authors approved the final version.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.A. (adriano.aguzzi@usz.ch).

Reviewer Information *Nature* thanks D. Pleasure and the other anonymous reviewer(s) for their contribution to the peer review of this work.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

Mice. Mice were bred in specified-pathogen-free facilities at the University Hospital Zurich and Washington University, and housed in groups of 3–5, under a 12 h light/12 h dark cycle (from 7 a.m. to 7 p.m.) at 21 ± 1 °C, with sterilized chow food (Kliba No. 3431, Provimi Kliba) and water *ad libitum*. Animal care and experimental protocols were in accordance with the Swiss Animal Protection Law, and approved by the Veterinary Office of the Canton of Zurich (permits 123, 130/2008, 41/2012 and 90/2013). The following mice were used in the present study: C57BL/6J, $Prnp^{ZH1/ZH1}$ (ref. 3), co-isogenic C57BL/6J $Prnp^{ZH3/ZH3}$ and $Prnp^{WT/WT}$ control mice[6] and Schwann cell-specifc Dhh^{Cre}::$Gpr126^{fl/fl}$ mutants[3,4]. Mice of both genders were used for experiments unless specified. Archival tissues from previous studies[1,6] were also analysed in the current study.

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment except where stated.

Primary Schwann cell culture. Sciatic nerves from postnatal day 2–5 were dissected using microsurgical techniques. Nerves were dissociated in serum-free DMEM supplemented with 0.05% collagenase IV (Worthington) for 1 h in the incubator. Sciatic nerves were mechanically dissociated using fire-polished Pasteur pipettes. Cells were filtered in a 40-μM cell strainer and washed in Schwann cell culture medium (DMEM, Pen-Strep, Glutamax, FBS 10%) by centrifugation at 300*g* for 10 min. Resuspended cells were plated on 3.5 cm Petri dishes previously coated with poly-L-lysine 0.01% (w/v) and laminin (1 mg/ml). Laminin (Cat. No: L2020; from Engelbreth-Holm-Swarm murine sarcoma basement membrane) and poly- L-lysine were obtained from Sigma-Aldrich.

Purification of recombinant proteins. Full-length recombinant PrP (recPrP, residues 23–231) and globular domain (GD, residues 121–231) were purified as previously described[21–23]. The generation of the GST fusion FT-PrP expression vector (pGEX-KG FT-PrP) was described previously; a modified purification protocol was used[24]. The FT-PrP expression vector was transformed into BL21 (DE3) strain of *Escherichia coli* (Invitrogen). Bacteria were grown in Luria-Bertani medium to an OD of 0.6, and the expression of the fusion protein was induced with 0.5 mM isopropyl-1-thio-β-D-galactopyranoside (AppliChem). Cells were then grown for another 4 h at 37 °C and 100 rpm shaking. Cells were pelleted at 5,000*g* for 20 min at 4 °C (Sorvall centrifuge, DuPont). The pellet was resuspended on ice in lysis buffer (phosphate-buffered saline supplemented with complete protease inhibitors (EDTA-free, Roche), phenylmethyl sulfonyl fluoride (Sigma) and 150 μM lysozyme (Sigma)) and incubated on ice for 30 min. Triton-X 100 (1%), $MgCl_2$ (10 mM) and DNase I (5 μg/ml, Roche) were added, and the lysate was incubated on ice for 30 min. The lysate was than centrifuged for 20 min at 10,000*g* at 4 °C. Glutathione sepharose beads were washed with PBS and incubated with the cell lysate for 1 h at 4 °C on a rotating device. Beads were packed into a column and washed with PBS until a stable baseline was reached as monitored by absorbance at A_{280} using an ÄKTAprime (GE healthcare). The fusion protein was cleaved on the beads with 5 U/ml Thrombin (GE Healthcare) for 1 h at room temperature under agitation. For thrombin removal, benzamidine sepharose beads were added and incubated for 1 h at 4 °C on a rotating wheel. Protein preparations were analysed by 12% NuPAGE gels followed by Coomassie- or silver-staining. To achieve a higher purity of the protein, we next applied the protein to a sulfopropyl (SP) sepharose column equilibrated with 50 mM Tris-HCl buffer, pH 8.5. Elution was performed with a linear NaCl gradient of 0–1,000 mM. Fractions containing the protein were collected and concentrated (AMICON; MWCO 3500). The protein was then injected in 500 μl portions into a size-exclusion chromatography system (TSK-GEL $G2000SW_{XL}$ column (Tosoh Bioscience)) and eluted with a linear gradient using PBS. Pure fractions were combined, concentrated and stored at −20 °C. The purity of FT-PrP was >95–98% as judged by a silver-stained 12% NuPAGE gel.

Cell culture. SW10 cells and clones derived from them were all grown in DMEM medium supplemented with 10% fetal bovine serum (FBS), penicillin-streptomycin and Glutamax (all obtained from Invitrogen). HEK293T cells, its clonal variant HEK293(H) cells and clones derived therefrom overexpressing various GPCRs were grown in DMEM-F12 medium supplemented with 10% FCS, penicillin-streptomycin and Glutamax (all obtained from Invitrogen). All cell lines were regularly monitored for mycoplasma contamination. The authenticity of SW10 and its derivatives was established by monitoring the expression of Schwann-cell specific markers (Extended Data Fig. 6a). Human Gpr126 (NM_020455), Gpr124, Gpr64, Gpr56, Gpr133, Gpr56 and Gpr176 expression plasmids (pCGpr126-V5, pCGpr124-V5, pCGpr65-V5, pCGpr56-V5, pCGpr133-V5, pCGpr56-V5 and pCGpr176-V5) were generated by PCR amplification of the respective cDNA followed by TOPO cloning into the pCDNA3.1/V5-His-TOPO vector. The cDNA was in frame with the V5 tag (sequence: GKPIPNPLLGLDST) at the C terminus. HEK^{GPR126} and HEK^{GPR176} cells were generated by transfecting 1 μg of plasmid into one well of a subconfluent 6-well plate using 3 μl Fugene (Roche) according to the manufacturer's protocol. Twenty-four hours after transfection, cells were transferred to a 10-cm dish and grown in selective medium containing 0.4 mg/ml G418 (Invitrogen) until emergence of resistant colonies. A limiting dilution was carried out to obtain clonal lines. Membrane expression of the transgene was assessed in the selected clones by confocal microscopy using 1:100 diluted anti-V5 antibody (Invitrogen) and the Cytofix/Cytoperm kit (Pharmingen Cat. Nr. 554714), according to the manufacturer's protocol.

Cerebellar granule neuronal culture. Cerebellar granule neurons were generated from 7–8-day-old $Prnp^{ZH1/ZH1}$ mice as described previously[25]. Cultures were plated at 350,000 cells per cm^2 in Basal Medium Eagle (BME) (Invitrogen) with 10% (v/v) FCS and maintained at 37 °C in 5% CO_2.

Plasmids and transfections. pCDNA-PrP^C was generated by cloning murine PrP^C into pCDNA3.1 vector as described previously[26]. A site-specific mutagenesis kit (Stratagene) was used to induce alanine substitutions of QPSPG and KKRPK domains in PrP^C. Primers used for generating the Ala-QPSPG plasmid were: forward, GTG GAA GCC GGT ATC CCG GGG CGG CAG CCG CTG CAG GCA ACC GTT ACC C; reverse, GGG TAA CGG TTG CCT GCA GCG GCT GCC GCC CGG GGA TAC CGG CTT CCA C. Primers for Ala-KKRPK were: forward, CTA TGT GGA CTG ATG TCG GCC TCT GCG CAG CGG CGC CAG CGC CTG GAG GGT GGA ACA CCG; reverse, CGG TGT TCC ACC CTC CAG GCG CTG GCG CCG CTG CGC AGA GGC CGA CAT CAG TCC ACA TAG. Transfections were performed with Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol. 3 μg of DNA was used per well of a 6-well plate. Cells were washed 24 h after transfection using PBS, and fresh medium was added to the cells.

Immunoprecipitation. HEK293T and HEK^{GPR126} cells growing in T75 flasks at 50% density were treated with recombinant FT or GD (2 μM, 20 min). Cells were washed twice in PBS and lysed in IP buffer: 1% Triton X-100 in PBS, 1× protease inhibitors (Roche) and Phospho stop (Roche) for 20 min on ice followed by centrifugation at 5000 rpm for 5 min at 4 °C. BCA assays were performed to quantify the amount of protein, and 500 μg of protein was used for immunoprecipitations. 2 μg anti-V5 antibody was added to the cell lysate and incubated on a wheel rotator overnight at 4 °C. On the following day, Protein G dynabeads (Invitrogen) were added to the samples and incubated for a further 3 h on the wheel at 4 °C. Beads were washed three times for 5 min each using the IP buffer followed by addition of 2× sample buffer containing DTT (1 mM final). Samples were heated at 95 °C for 5 min, loaded on 4–12% Novex Bis-tris gels (Invitrogen), and migrated for 1.5 h at 150 V followed by western blotting. Immunoprecipitations were performed by adding 2 μg of POM2 antibody to 500 μl of cell medium and incubating overnight on a wheel rotator at 4 °C. Protein G beads were then added, and incubation on a wheel rotator at 4 °C was performed again.

RNA isolation and quantitative PCR. RNA extraction and quantitative PCR were performed as described previously[1]. The following primers were used: EGR2 forward: 5′-AATGGCTTGGGACTGACTTG-3′; EGR2 reverse: 5′-GCCAGAGAAACCTCCATT-3′; GAPDH forward: 5′-CCACCCCAGCA AGGAGAC-3′; GAPDH reverse: 5′-GAAATTGTGAGGGAGATGCT-3′.

Zebrafish mutant strains and analysis. Adult zebrafish were maintained in the Washington University Zebrafish Consortium facility (http://zebrafishfacility.wustl.edu/) and all experiments were performed in compliance with institutional protocols. Embryos were collected from harem matings or *in vitro* fertilization, raised at 28.5 °C, and staged according to standard protocols[27]. The $gpr126^{st49}$ and $gpr126^{st63}$ mutants were described previously[7,8].

Zebrafish peptide treatment, immunostaining, and quantification. $gpr126^{st63}$ or $gpr126^{st49}$ mutants were collected from homozygous mutant crosses and wild-type larvae were collected from AB* strain crosses and raised to 50 hpf. $FT_{23–50}$ treatment of *gpr126* mutants was performed as previously described[15]. Briefly, egg water was replaced with either 20 μM $FT_{23–50}$ in egg water or egg water containing an equivalent volume of DMSO. At 55 hpf, larvae were washed twice and raised in egg water to 5 dpf. Wild-type and *gpr126* larvae were fixed in 2% paraformaldehyde plus 1% tricholoroacetic acid in phosphate buffered saline, and Mbp and acetylated tubulin immunostaining was performed as described previously[8,28]. Expression scoring was performed with observers blinded to treatment according to the following rubric: strong, strong and consistent expression throughout PLLn; some, weak but consistent expression in PLLn; weak, weak and patchy expression in PLLn; none, no expression in PLLn. n = three independent replicate $gpr126^{st63}$ assays and one $gpr126^{st49}$ assay. n = 87 DMSO-treated $gpr126^{st63}$ larvae, 81 Prp-FT-treated $gpr126^{st63}$ larvae, 27 DMSO-treated $gpr126^{st49}$ larvae, 25 Prp-FT-treated $gpr126^{st49}$ larvae.

Fluorescent nerve images were analysed using the Fiji software[29]. A rectangular region-of-interest (ROI) was drawn longitudinally over the fluorescent nerve. The longitudinal grey-scale histogram of the myelin basic protein (Mbp) was normalized pixel-by-pixel to the corresponding intensity of the acetylated tubulin (AcTub). The size of the measured ROIs was kept constant across different treatment modalities.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Flow cytometry. SW10 cells were grown in P75 flasks at 50% density, rinsed with PBS, and detached from culture flasks with dissociation buffer containing EDTA (GIBCO). After detaching, cells were washed to remove residual EDTA and counted using a Neubauer chamber. Batches of 10^5 SW10 cells were transferred to FACS tubes, treated with HA-tagged recombinant peptides for 20 min, washed, and incubated with Alexa-488 conjugated anti-HA antibody for 30 min. After further washes and centrifugations, cells were resuspended in 200 µl FACS buffer (PBS + 10% FBS) and analysed with a FACS Canto II cytofluorimeter (BD Biosciences). Data were analysed using FloJo software.

Western blot analysis. Schwann cells were lysed in cell-lysis buffer (Tris-HCl 20 mM, NaCl 137 mM, Triton-X-100 1%) supplemented with protease inhibitor cocktail (Roche complete mini). The lysate was homogenized by passing several times through a 26G syringe, and cleared by centrifugation at 8,000*g*, 4 °C for 2 min. in a tabletop centrifuge (Eppendorf 5415R). Protein concentration was measured with the BCA assay (Thermo Scientific). 10 µg total protein was boiled in 4 × LDS (Invitrogen) at 95 °C for 5 min. After a short centrifugation, samples were loaded on a gradient of 4–12% Novex Bis-Tris Gel (Invitrogen) for electrophoresis at constant voltage of 200 V. Gels were transferred to PVDF membranes with the iBlot system (Life technologies). Membranes were blocked with 5% Top-Block (Sigma) in PBS-T for 1h at room temperature. Primary antibody was incubated overnight in PBS-T with 5% Top-Block. Membranes were washed three times with PBS-T for 10 min and incubated for 1 h with secondary antibodies coupled to horseradish peroxidase at room temperature. After three washes with PBS-T, the membranes were developed with a Crescendo chemiluminescence substrate system (Millipore). Signals were detected using a Stella 3200 imaging system (Raytest).

Antibodies. Monoclonal antibodies against PrP^C were obtained and used as described previously[4]. Fab3 and Fab71 antibodies were generated using the phage display technology and their epitopes were mapped with overlapping peptides. Anti AKT, p-AKT were obtained from Cell signaling and used at 1:2,000 dilutions for western blotting. The anti-p75NGF receptor antibody was obtained from Abcam and used at a 1:200 dilution for immunofluorescence. Anti V5 antibody was from Invitrogen and used at a dilution of 1:500 for western blot and 2 µg antibody was used for immunoprecipitation on 500 µg of cell lysate.

Cyclic AMP measurements. In the direct cAMP ELISA assay, cAMP levels were assessed with a colorimetric competitive immunoassay (Enzo Life Sciences). Quantitative determination of intracellular cAMP was performed in cells or tissues lysed in 0.1 M HCl to stop endogenous phosphodiesterase activity and to stabilize the released cAMP. SW10 or HEK293T cells (100,000 cells per well) were plated in 6-well plates to ~50% density. Cells were treated with conditioned medium or recombinant peptides (2 µM, unless specified) for 20 min unless otherwise mentioned. Cells were lysed with 0.1 M HCl lysis buffer (Direct cAMP ELISA kit, Enzo). To ensure complete detachment of cells, cell scrapers were used. Lysates were homogenized with a 26G needle and syringe before clearing by centrifugation at 600*g* for 10 min. The subsequent steps were performed according to the manufacturer's protocol based on competition of sample cAMP with a cAMP-alkaline phosphatase conjugate. To measure *in vivo* cAMP changes, BL6, $Prnp^{ZH3/ZH3}$ or $Prnp^{ZH1/ZH1}$ mice were intravenously injected with 600 µg of either FT_{23-50} or, as a control, uncharged $FT_{23\text{-}50}$ ($FT^{K\rightarrow A}_{23-50}$). Twenty minutes after infusion, mice were killed and all organs were collected. For cAMP assays, organs were homogenized in 0.1 M HCl. Subsequent steps were performed according to the manufacturer's protocols as described above. Cyclic AMP levels were calculated using a cAMP standard curve in the case of ELISA based assay. Finally, cAMP concentrations were normalized to total protein content in each sample. cAMP changes are represented as fold changes to the respective controls. For each experiment, at least three independent biological replicates were used. For *in vivo* assays, groups of 8–16 mice were used for each experiment. For normalization purposes, the median value of the respective control sample was defined as 1. All measurements within each panel were normalized to this control value. For *in vivo* assays, sample sets were coded and investigators were blinded to their identities. The assignment of codes to sample identities was performed only after the cAMP values were plotted for each set.

Generation of Schwann cell lines devoid of Gpr126. We designed two CRISPR short-guide RNA (sgRNAs) against exon 2 of Gpr126 (upper Guide CCTGTGTTCCTCTCTCAGGT and lower Guide AACAGGAACAGCAGG GCGCT). The DNA sequences corresponding to the sgRNAs were cloned into expression plasmids and transfected with EGFP-expressing Cas9-nickase plasmids. Single EGFP-expressing Schwann cells were isolated with a FACS sorter (Aria III). To determine the exact sequence of indels induced by genome editing, we amplified the sgRNA-targeted locus by PCR and subcloned the fragments into blunt-TOPO vectors. Ten colonies per cell line were sequenced and showed distinct indels on each allele. A clonal subline devoid of Gpr126 was used for further studies. This cell line possessed insertions on both the alleles; a 49-bp insertion at position 118 and a 5-bp insertion at position 84 on each allele. Both insertions led to a frameshift and to the generation of premature stop codons leading to early translation termination.

Promoter luciferase assay. Luciferase reporter constructs were generated containing a 1.3-kB sequence upstream of the transcription-starting site of Egr2. SW10 Schwann cells were transfected with Egr2 reporter construct and a renilla plasmid using lipofectamine 2000. After one day *in vitro*, Schwann cells were treated with recombinant full-length PrP (23–231), the globular domain of PrP (121–231) or PBS control. Luciferase activity was measured 24 h after stimulation with Dual-Luciferase Reporter Assay System (Promega) according to the manufacturer's recommendations. Results were normalized to renilla transfection controls.

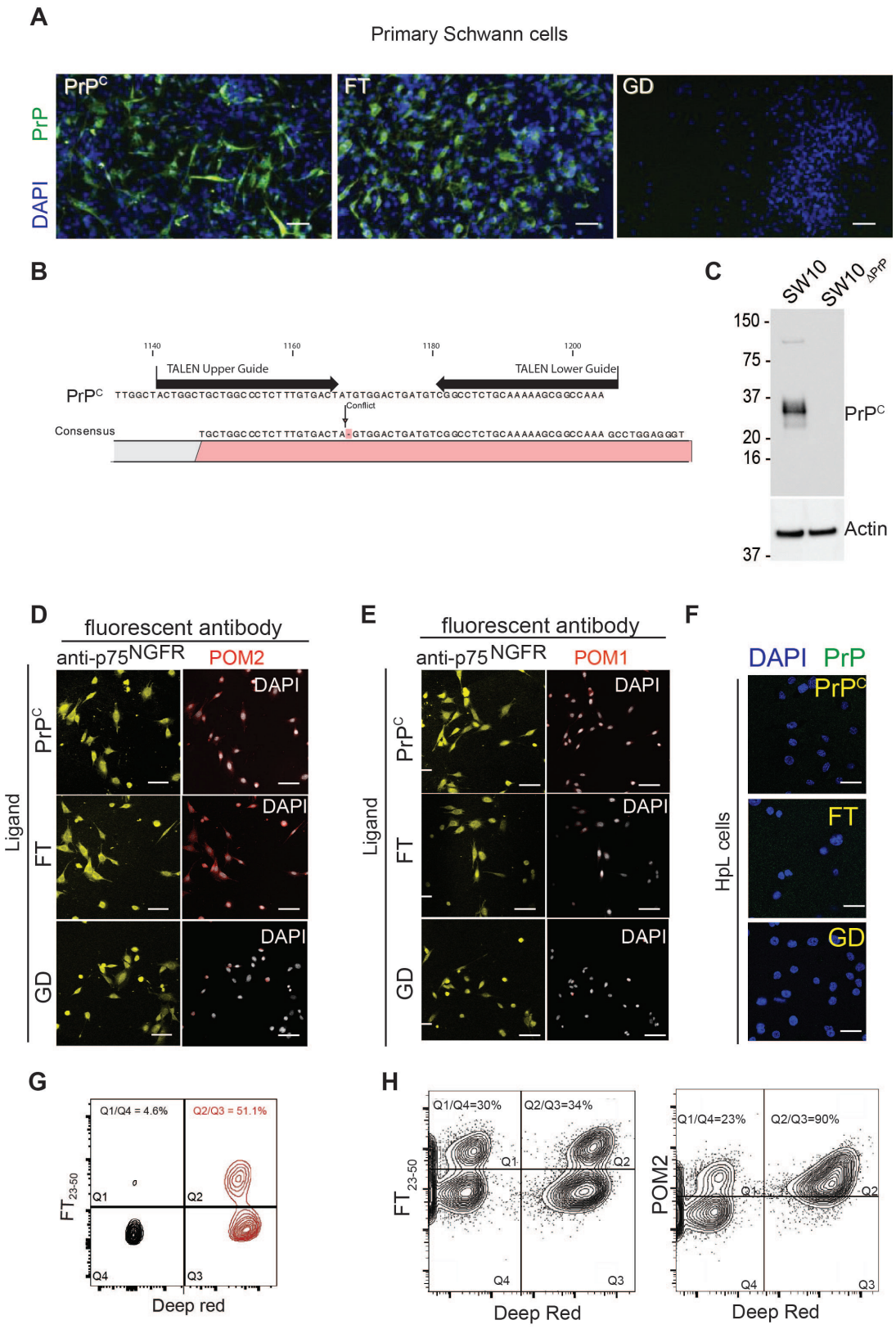
Immunocytochemistry. Glass coverslips were placed in 12-well plates (Thermo Scientific) and coated with 0.01% w/v Poly-L-lysine solution (Sigma) overnight at room temperature. Coverslips were washed three times with ddH_2O and dried for 2 h in a laminar-flow hood. Schwann cells were seeded and cultured at 50% density. Cells were treated with recombinant FT-PrP, full length recPrP or C1-PrP for 20 min, and washed with serum-free DMEM. Cells were further washed with PBS followed by fixation with 4% paraformaldehyde. Fixed cells were incubated in blocking buffer (PBS+10% FBS) for 1 h. Cells were treated with various primary antibodies followed by washes and incubation with Alexa 488 and Alexa 647 tagged rabbit or mouse secondary antibodies (Life Technologies). Imaging was performed by Leica SP2 confocal microscope using a 20× objective; images were processed by Image J software.

Transmission electron microscopy. Transmission electron microscopy was performed as previously described[6]. Briefly, mice under deep anaesthesia were subjected to transcardial perfusion with PBS heparin and sciatic nerves were fixed *in situ* with 2.5% glutaraldehyde plus 2% paraformaldehyde in 0.1 M phosphate buffer, pH 7.4 and embedded in Epon. Ultrathin sections were mounted on copper grids coated with Formvar membrane and contrasted with uranyl acetate/lead citrate. Micrographs were acquired using a Hitachi H-7650 electron microscope (Hitachi High-Tech, Japan) operating at 80 kV. Brightness and contrast were adjusted using Photoshop. For quantification of Remak bundles and onion bulb-like structures, images were captured at 1,500× magnification and axon numbers and abnormal onion bulb-like structures were counted manually. Quantification was performed in a blinded fashion by assigning numbers to the images and upon completion of quantification genotypes were revealed.

Recombinant peptides. HA-tagged and untagged synthetic peptides were produced by EZ Biosciences. A stock solution of 2 mM was prepared by dissolving the peptides in PBS and they were used at a final concentration of 2 µM unless specified. The sequences of all the peptides used in this study can be found in Extended Data Table 1.

21. Zahn, R., von Schroetter, C. & Wüthrich, K. Human prion proteins expressed in *Escherichia coli* and purified by high-affinity column refolding. *FEBS Lett.* **417,** 400–404 (1997).
22. Lysek, D. A. & Wüthrich, K. Prion protein interaction with the C-terminal SH3 domain of Grb2 studied using NMR and optical spectroscopy. *Biochemistry* **43,** 10393–10399 (2004).
23. Hornemann, S., Christen, B., von Schroetter, C., Pérez, D. R. & Wüthrich, K. Prion protein library of recombinant constructs for structural biology. *FEBS J.* **276,** 2359–2367 (2009).
24. Guillot-Sestier, M. V., Sunyach, C., Druon, C., Scarzello, S. & Checler, F. The α-secretase-derived N-terminal product of cellular prion, N1, displays neuroprotective function in vitro and in vivo. *J. Biol. Chem.* **284,** 35973–35986 (2009).
25. Schulz, J. B., Weller, M. & Klockgether, T. Potassium deprivation-induced apoptosis of cerebellar granule neurons: a sequential requirement for new mRNA and protein synthesis, ICE-like protease activity, and reactive oxygen species. *J. Neurosci.* **16,** 4696–4706 (1996).
26. Dametto, P. *et al.* Neurodegeneration and unfolded-protein response in mice expressing a membrane-tethered flexible tail of PrP. *PLoS One* **10,** e0117412 (2015).
27. Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B. & Schilling, T. F. Stages of embryonic development of the zebrafish. *Dev. Dyn.* **203,** 253–310 (1995).
28. Lyons, D. A. *et al.* erbb3 and erbb2 are essential for Schwann cell migration and myelination in zebrafish. *Curr. Biol.* **15,** 513–524 (2005).
29. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9,** 676–682 (2012).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

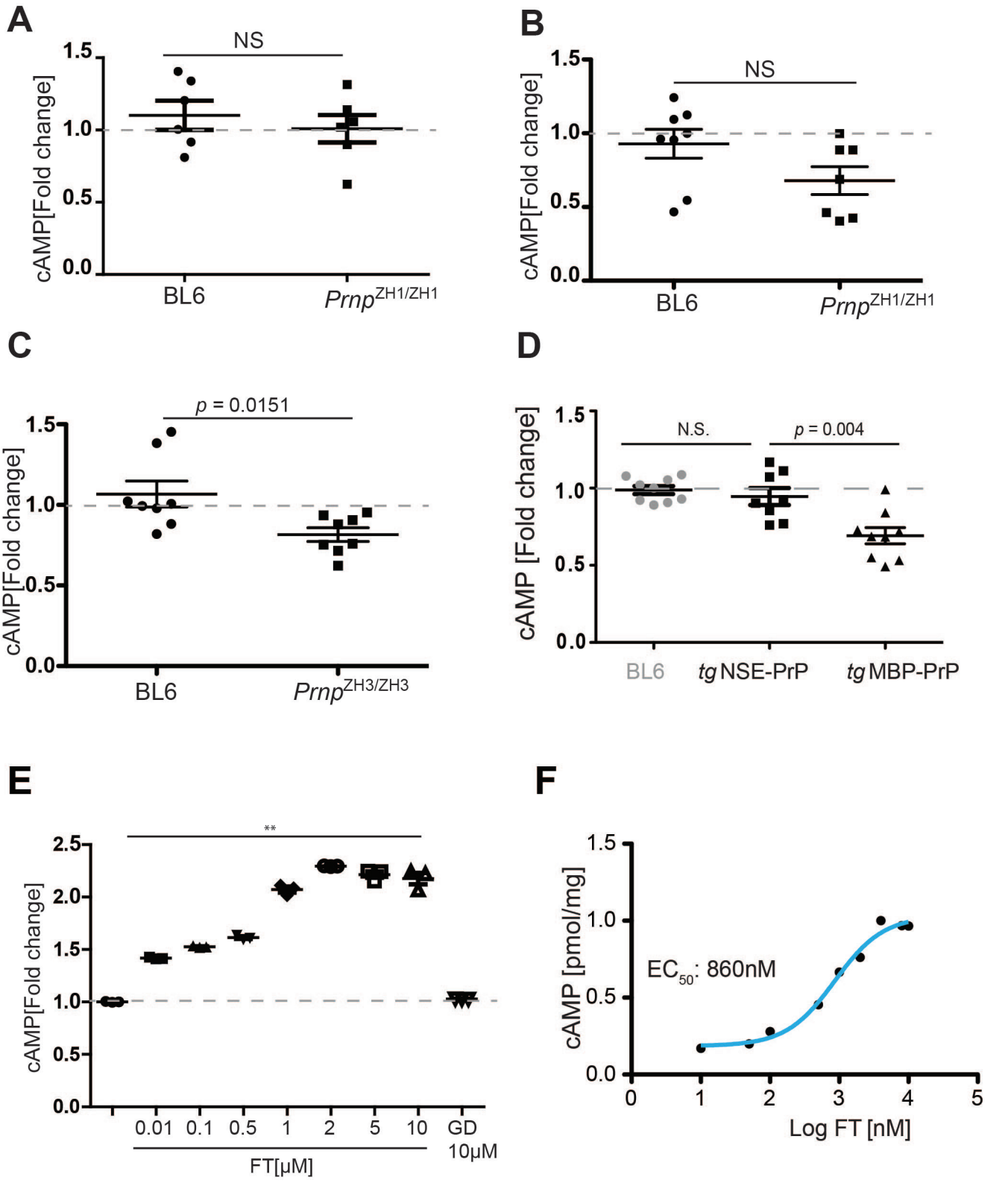


Extended Data Figure 1 | See next page for caption.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

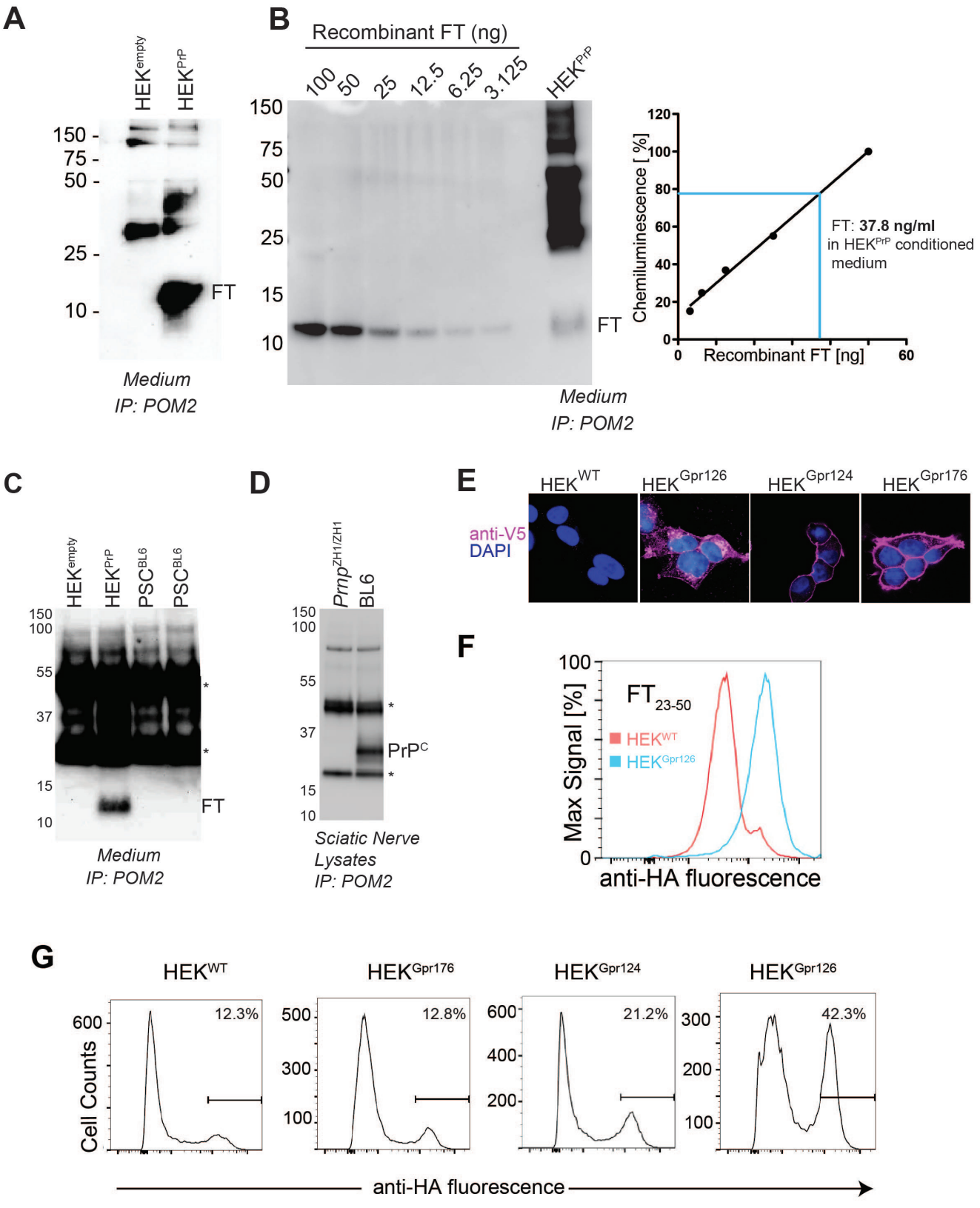
Extended Data Figure 1 | FT binds to Schwann cells through a proteinaceous receptor. a, Primary Schwann cells were isolated from the sciatic nerves of *Prnp*$^{ZH1/ZH1}$ mice and grown on coverslips. Cells were exposed for 20 min to recombinant PrP^C, FT, or GD (2 μM), fixed, and stained with POM2 (FT, PrP^C) or POM1 (GD). Antibodies were visualized in the green channel and nuclei were stained with DAPI (blue). PrP^C and FT, but not GD, adhered to the cells. Scale bars, 25 μm. **b**, Schematic representation of the target region for transcription activator-like endonucleases (TALEN) in the *Prnp* gene. Target guides are indicated by arrows. Gene editing resulted in a deletion leading to a frame shift in the PrP^C coding sequence (designated as a conflict in the figure) and a premature stop codon identified by sequencing. **c**, Wild-type SW10 cells and a subclone isolated after treatment with TALEN ($SW10_{\Delta PrP}$) were probed by western blotting using POM1. $SW10_{\Delta PrP}$ showed complete abrogation of PrP^C expression and was used for further experiments. Levels of actin on the same membrane were monitored to confirm equal loading of cell lysates onto the gel. For uncropped gels see Supplementary Information File 1. **d**, **e**, $SW10_{\Delta PrP}$ cells were treated with full-length recombinant (PrP^C, residues 23–231), flexible tail (FT, 23–110), or globular domain (GD, 121–231). PrP epitopes were detected with POM2 (**d**) or POM1 (**e**, red). Grey, DAPI. As expected, FT was detected only by POM2. Cells were also labelled with antibodies to the p75 nerve-growth factor receptor (yellow), a Schwann cell marker. PrP^C and FT, but not GD, adhered to Schwann cells. Scale bar, 26 μm. **f**, The PrP^C-deficient cell line HpL^5 was treated with recombinant PrP^C, FT, and GD as in **a**. None of the recombinant proteins adhered to HpL cells. Scale bars, 20 μm. **g**, SW10 cells were trypsinized, washed, and mixed with non-trypsinized SW10 cells labelled with Deep Red cell tracker. Cells were incubated with HA-tagged peptide $FT_{23–50}$, and binding was visualized by flow cytometry. The Deep Red signal (abscissa) was used to differentiate trypsinized from non-trypsinized cells. 51% of untreated cells, but only 5% of trypsinized cells, became decorated by $FT_{23–50}$–HA, indicating that $FT_{23–50}$ reacted with trypsin-sensitive surface molecules. **h**, SW10 cells were digested (30 min) with phosphatidylinositol phospholipase C (PI-PLC, 0.5 U), washed, and incubated with $FT_{23–50}$–HA along with undigested Deep Red-labelled cells (left). The proportion of binders in the digested (34%) and undigested samples (30%) was similar, indicating that the $FT_{23–50}$ receptor was neither PrP^C itself nor any other GPI-linked protein. To monitor the efficiency of PI-PLC treatment, we assessed POM2 binding to PrP^C on both treated and untreated cells (right). POM2 binding was significantly decreased in PI-PLC treated cells (23%) compared to untreated cells (90%). Panels depict biologically independent triplicates; unpaired Student's *t*-test was used for statistical analysis.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 2 | *Prnp* ablation results in lower cAMP levels in sciatic nerves. a, **b**, cAMP was measured in sciatic nerves isolated from 4-day-old (**a**) or 4-week-old (**b**) BL6 and *Prnp*$^{ZH1/ZH1}$ mice. No difference was observed in cAMP levels in 4-day-old mice, whereas 4-week-old *Prnp*$^{ZH1/ZH1}$ mice displayed a trend towards decreased cAMP levels. **c**, Sciatic nerves from 10-week-old *Prnp*$^{ZH3/ZH3}$ mice showed a significant decrease in cAMP ($P = 0.0151$). **d**, PrP^{C} expression by neurons (*tg*NSE-PrP), but not by Schwann cells (*tg*MBP-PrP), restored cAMP levels in sciatic nerves of 10–16-week-old mice. Each dot represents one mouse. **e**, SW10 cells were seeded in 6-well plates and treated (20 min) with recombinant FT or GD (10 μM). Addition of FT, but not of GD, resulted in a concentration-dependent intracellular cAMP increase. **f**, Primary *Prnp*$^{ZH1/ZH1}$ Schwann cells were treated with FT (20 min), and cAMP concentrations were determined by immunoassay. A dose–response curve was interpolated. Data are representative of three biologically independent experiments and statistical significance was evaluated by unpaired Student's *t*-tests. Error bars show s.e.m.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

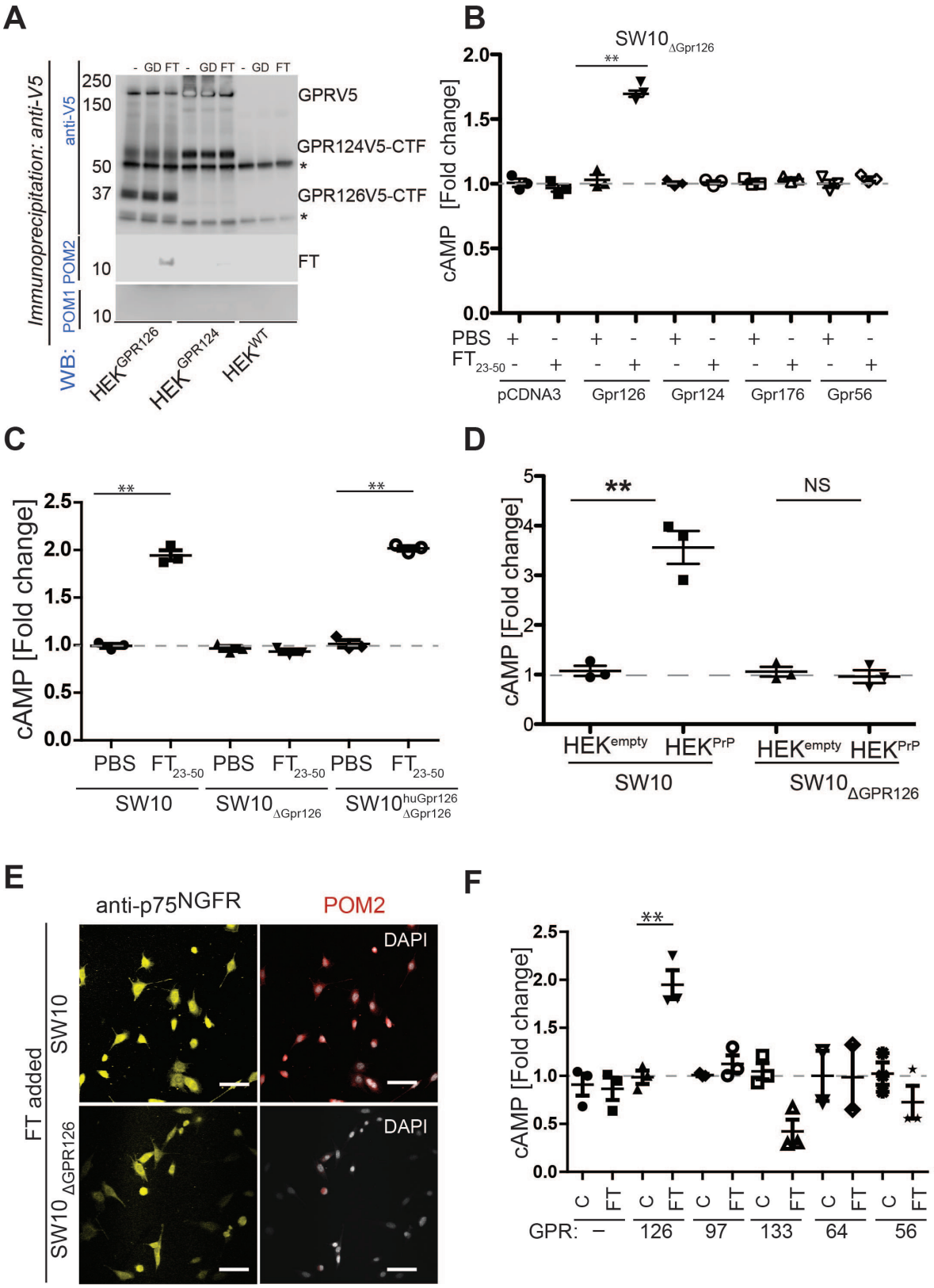


Extended Data Figure 3 | See next page for caption.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Figure 3 | FT proteolytically released from full-length PrP^C binds to Gpr126. a, HEK293T cells were transfected with an empty vector (HEK^{empty}) or a plasmid expressing murine PrP^C (HEK^{PrP}). Cell medium was collected 48 h after transfection and subjected to immunoprecipitation with monoclonal antibody POM2 (against PrP^C), followed by western blotting using biotinylated POM2 and streptavidin-HRP. FT was observed only in the medium from HEK^{PrP} cells. **b**, FT released into the conditioned medium of HEK^{PrP} was immunoprecipitated using POM2 and visualized by western blotting with biotinylated POM2. Various amounts of recombinant FT (3.125–100 ng) were used for calibration, and the concentration of FT released into 1 ml of the medium upon immunoprecipitation was estimated to be 37 ng ml^{-1}. **c**, Conditioned medium from primary BL6 Schwann cells cultures (PSC^{BL6}) was subjected to immunoprecipitation with antibody POM2 followed by western blotting with POM2. For control, we used conditioned medium from HEK cells transfected with a non-coding plasmid (HEK^{empty}) or a with a plasmid encoding murine PrP^C (HEK^{PrP}). FT was detected only in conditioned medium from HEK^{PrP} cells (lane 2) but not in conditioned medium from two independent PSC^{BL6} cultures (lanes 3 and 4). Asterisks denote immunoglobulins detected by the secondary antibody. **d**, Sciatic nerve lysates obtained from 10-week-old $Prnp^{ZH1/ZH1}$ and BL6 mice and subjected to immunoprecipitation with POM2 antibody followed by western blotting with POM2. Full-length PrP^C, but no FT, was detectable in the immunoprecipitates of nerves from BL6 mice. **e**, Wild-type HEK293T cells (HEK^{WT}) or HEK293T cells overexpressing various GPCRs bearing V5 epitope tags (HEK^{Gpr126}, HEK^{Gpr124}, and HEK^{Gpr176}) were grown on coverslips and stained with anti V5 antibody (detecting tagged GPCRs; magenta). Nuclei were stained with DAPI (blue). Staining revealed cell surface expression of all transfected GPCRs. Scale bar, 8 μm. **f**, HA-tagged $FT_{23–50}$ peptide (2 μM) was added to HEK^{WT} cells or to HEK^{GPR126} cells, labelled with anti-HA antibody, and subjected to cytofluorimetry. Overexpression of Gpr126 increased the binding of $FT_{23–50}$. **g**, Binding of HA-tagged $FT_{23–50}$ to HEK^{GPR126} cells (right, monitored by cytofluorimetry) was conspicuously increased over that of wild-type, Gpr176, and Gpr124-overexpressing HEK293T cells. Data are representative of three biologically independent experiments; statistical significance was evaluated by unpaired Student's *t*-test. Error bars show s.e.m.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

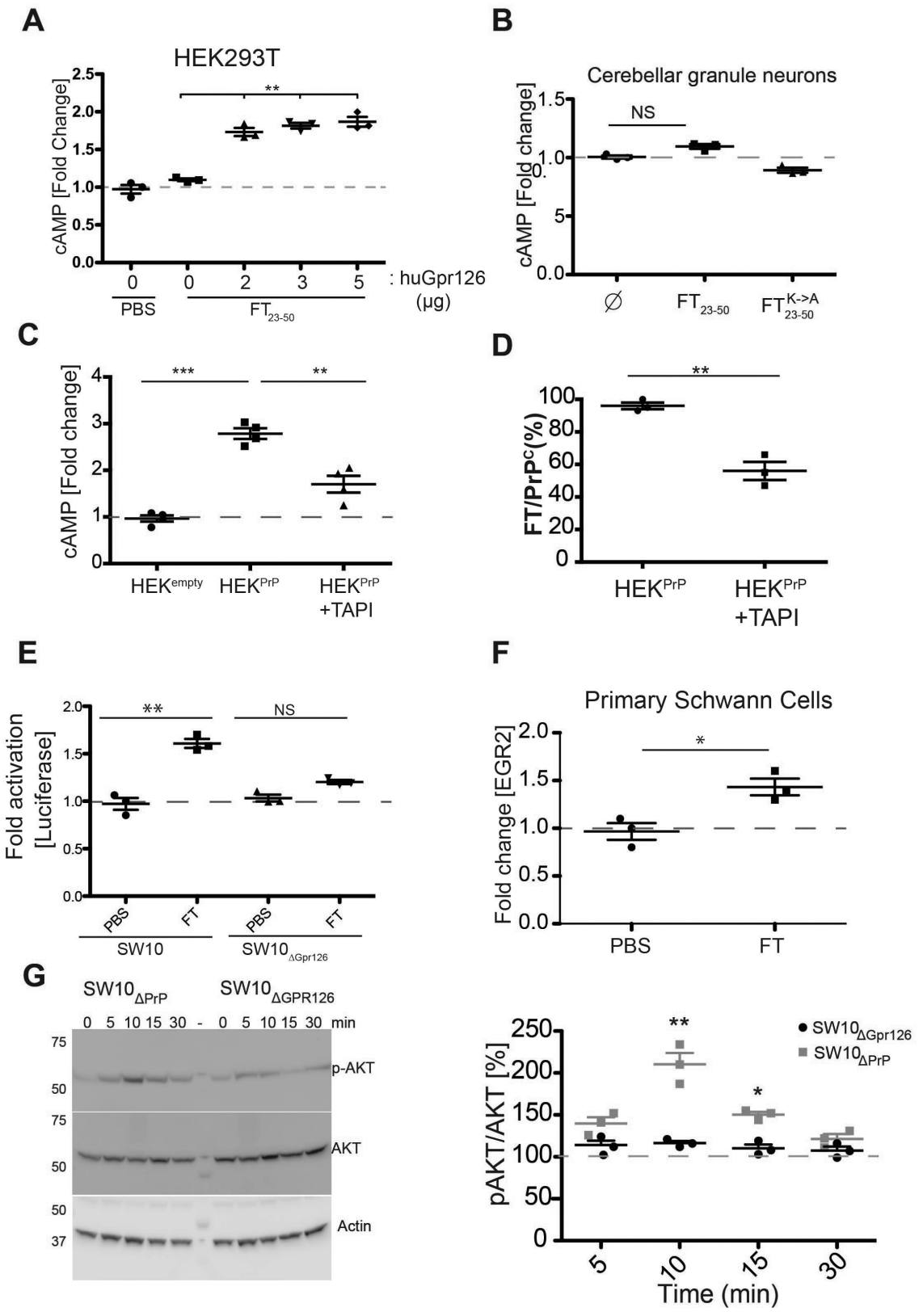


Extended Data Figure 4 | See next page for caption.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Figure 4 | FT binds selectively to Gpr126 and induces cAMP. a, HEK293T, HEK^{GPR124} and HEK^{GPR126} cells were exposed (20 min) to recombinant FT, GD (2 μM), or PBS, and subjected to immunoprecipitation using the anti-V5 antibody, followed by western blotting using POM2, anti-V5 or POM1. Anti-V5 detected full-length Gpr126, Gpr124 (denoted as GprV5 for both proteins) and the respective C-terminal fragments (Gpr126V5-CTF, Gpr124V5-CTF). POM2 revealed a band corresponding to the FT (lane 3) that co-precipitated with GPR126. POM1 indicated that GD did not bind. Lanes 1, 2 and 3: HEK^{GPR126} cells treated with PBS, GD and FT, respectively. Lanes 4, 5 and 6: HEK^{GPR124} cells treated with PBS, GD and FT, respectively. Lanes 7, 8 and 9: HEK293T cells treated with PBS, GD and FT, respectively. Asterisks: immunoglobulin heavy and light chains. For uncropped gels see Supplementary Information File 1. **b**, $SW10_{\Delta Gpr126}$ cells plated at a density of 100,000 cells per well in 6-well plates were transfected with control plasmid (pCDNA3) or plasmids encoding various GPCRs (Gpr126, 124, 176, and 56) bearing C-terminal V5 tags. Only cells transfected with pCGpr126-V5 showed a cAMP response to $FT_{23–50}$ 48 h post transfection. PBS treatment was used for control. **c**, Intracellular cAMP responses to FT treatment (2 μM, 20 min) in SW10 and $SW10_{\Delta Gpr126}$ cells, as well as $SW10^{huGpr126}_{\Delta Gpr126}$ cells expressing V5-tagged human Gpr126 (pCGpr126–V5). A significant increase in cAMP was observed in SW10 cells, whereas $SW10_{\Delta Gpr126}$ showed no change. In contrast, $SW10^{huGpr126}_{\Delta Gpr126}$ cells showed a significant cAMP increase, indicative of successful complementation. **d**, SW10 and $SW10_{\Delta Gpr126}$ cells were incubated (20 min) with conditioned medium from HEK^{empty} or HEK^{PrP} cells. HEK^{PrP}-conditioned medium induced a robust cAMP spike in SW10 but not $SW10_{\Delta Gpr126}$ cells. **e**, SW10 and $SW10_{\Delta Gpr126}$ cells were grown on coverslips for 24 h and exposed to recombinant FT (2 μM, 20 min). Cells were stained with POM2 (detecting FT, red; DAPI-stained nuclei, grey) and antibodies to $p75^{NGFR}$ (yellow). Deletion of Gpr126 largely suppressed FT binding. Scale bar, 26 μm. **f**, HEK293(H) cell lines were transfected with plasmids expressing different adhesion GPCRs (Gpr: 97, 133, 64, 56), followed by selection of cells expressing the receptor in presence of geneticin. GPCR-expressing cells and HEK^{Gpr126} cells were then treated with either $FT_{23–50}$ or $FT^{K\rightarrow A}_{23-50}$ for control (FT and C, respectively). Only cells expressing Gpr126 responded to $FT_{23–50}$ with a cAMP spike. Interestingly, cells expressing Gpr133 reacted with a decrease in cAMP levels. Data are representative of three biologically independent experiments; statistical significance was evaluated by unpaired Student's *t*-test. Error bars show s.e.m.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

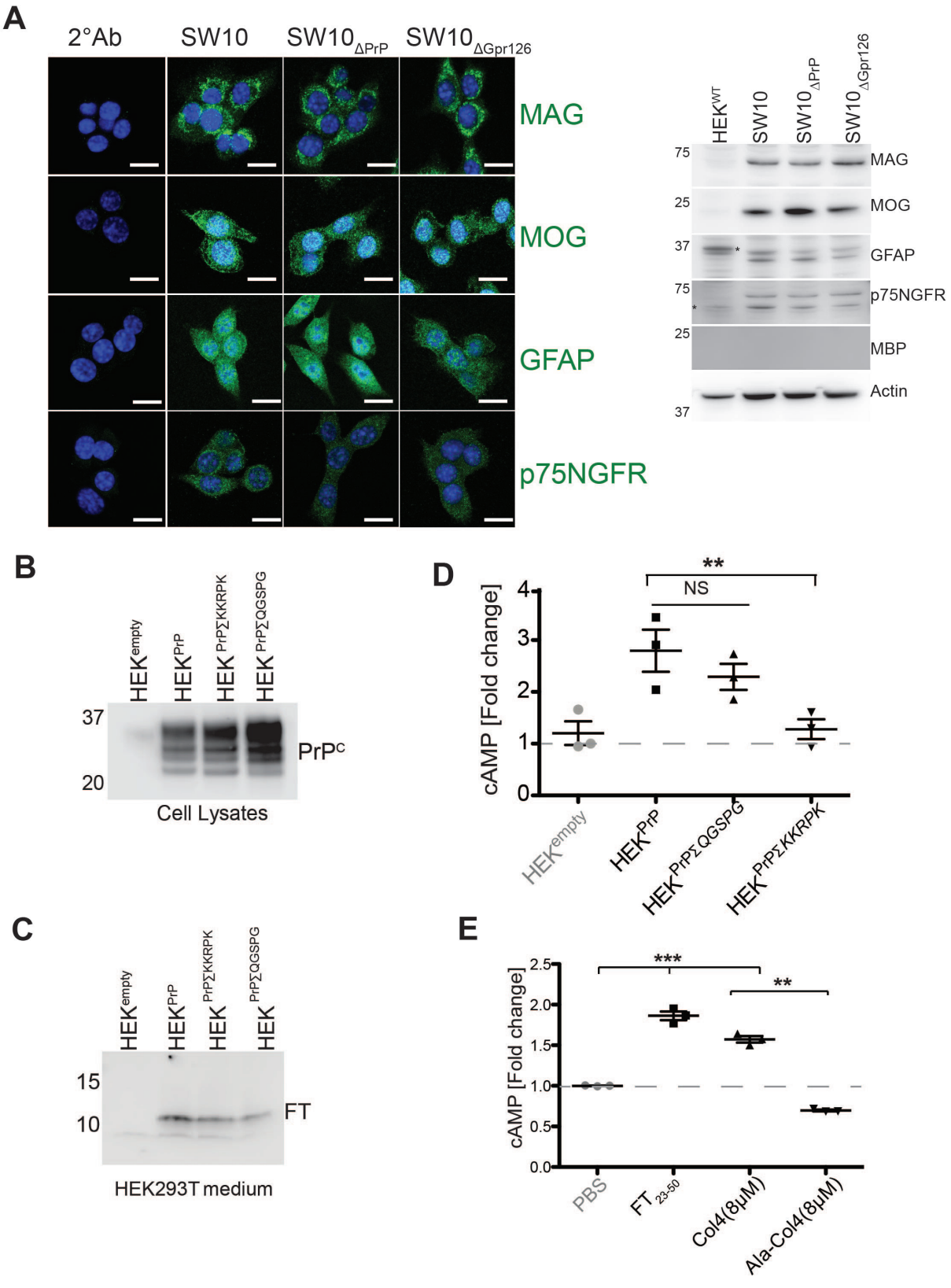


Extended Data Figure 5 | See next page for caption.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Figure 5 | FT promotes signal transduction in a Gpr126 dependent manner. a, HEK293T cells were transfected with increasing amounts of human Gpr126 plasmid (2–5 μg per well of a 6-well plate). 48 h post transfection, cells were treated with $FT_{23–50}$ or PBS as a control. Increasing amounts of Gpr126 cDNA did not result in amplification of the cAMP signal. **b**, Primary $Prnp^{ZH1/ZH1}$ cerebellar granule neuron cultures were seeded in 6-well plates at a density of 5×10^5 cells per well and treated with $FT_{23–50}$, $FT^{K\rightarrow A}_{23–50}$, or PBS. No alterations in the levels of cAMP were noticed. **c**, SW10 cells were exposed to conditioned medium from HEK cells that had been transfected with empty vector (HEK^{empty}) or a PrP^C expression vector (HEK^{PrP}). HEK^{PrP} were optionally treated with 100 μM of the TAPI-2 protease inhibitor for 24 h before harvesting the medium. TAPI-2 treatment resulted in reduced cAMP induction, suggesting that impaired proteolytic cleavage of the FT from PrP^C resulted in decreased signalling. **d**, Quantification of FT released into the medium relative to the total amount of PrP^C in lysates by western blotting. The spent medium of HEK^{PrP} cells treated with TAPI-2 contained less FT. **e**, SW10 and $SW10_{\Delta Gpr126}$ cells were transfected with an Egr2-controlled firefly luciferase reporter and treated with recombinant FT (2 μM) or PBS (24 h). Ordinate: luciferase expression normalized to a renilla luciferase control ($n = 3$; $^*P < 0.05$; *t*-test). Luciferase activity was observed only in SW10 cells stimulated with FT but not in $SW10_{\Delta Gpr126}$ cells. **f**, Primary Schwann cells were exposed to recombinant FT (2 μM, 1 h) or PBS. Egr2 mRNA expression was measured by quantitative RT–PCR and normalized against a panel of housekeeping genes. For uncropped gels, see Supplementary Information File 1. **g**, $SW10_{\Delta PrP}$ and $SW10_{\Delta Gpr126}$ cells were grown in 6-well plates, exposed to recombinant FT (≤30 min), and analysed by western blotting (left). Densitometry (right) showed increased phospho-AKT/AKT ratio in $SW10_{\Delta PrP}$ cells, but not in $SW10_{\Delta Gpr126}$ cells. Data are representative of three biologically independent experiments; statistical significance was evaluated by unpaired Student's *t*-test. Error bars show s.e.m.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

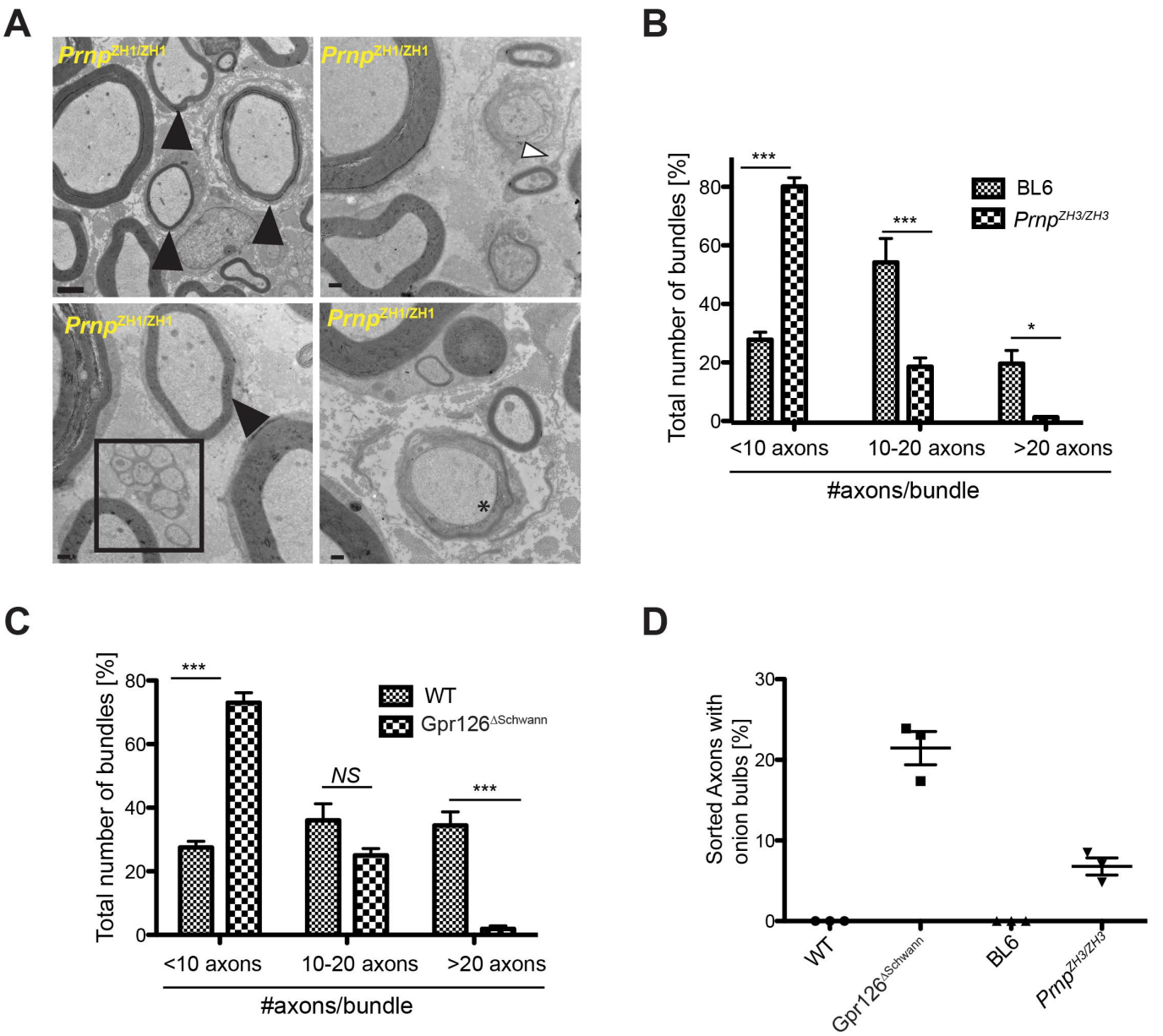


Extended Data Figure 6 | See next page for caption.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

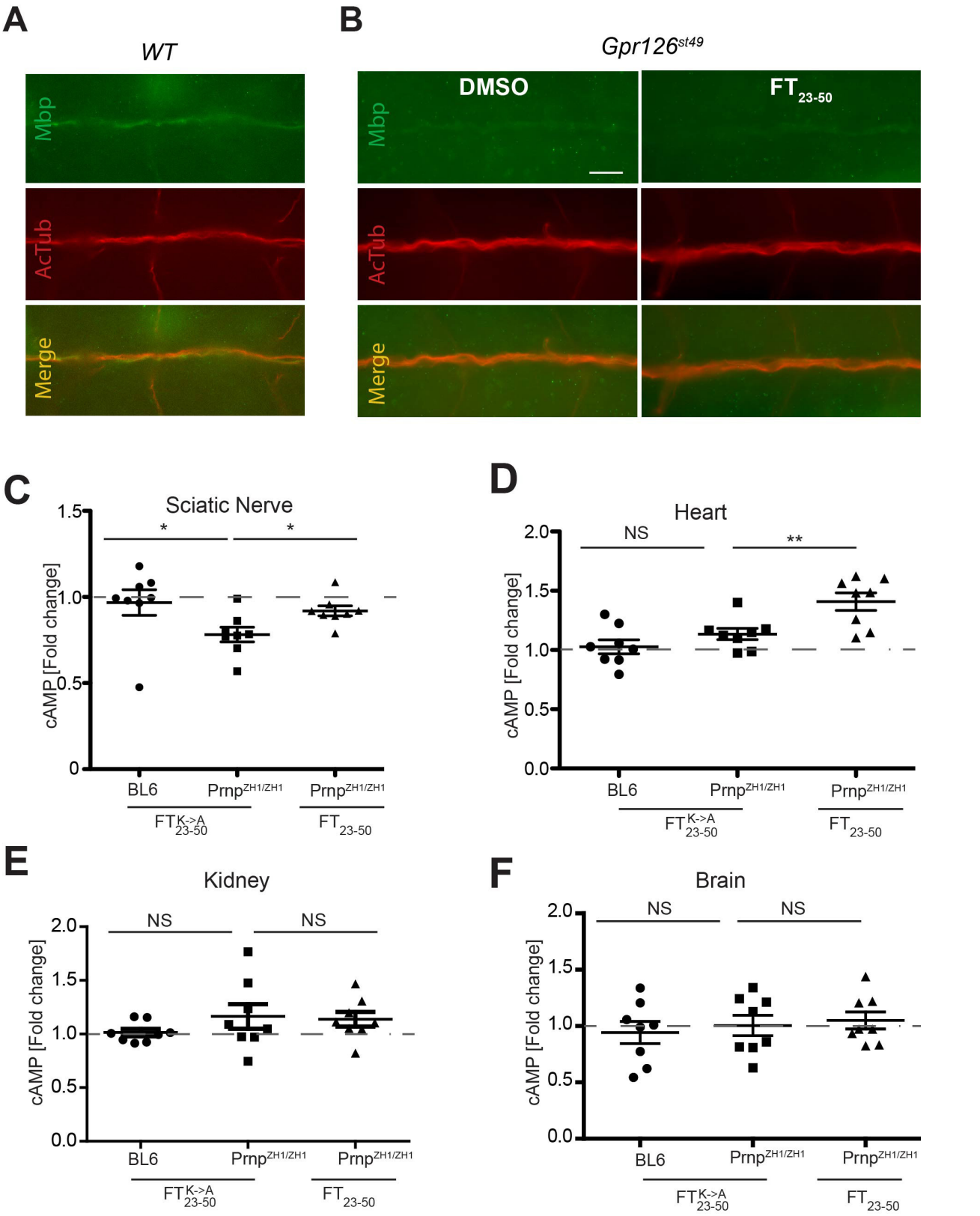
Extended Data Figure 6 | A domain conserved between FT and collagen-IV is required for cAMP induction. a, SW10, $SW10_{\Delta PrP}$ and $SW10_{\Delta Gpr126}$ cells were grown on coverslips and stained with antibodies against myelin-associated glycoprotein (MAG), myelin oligodendrocyte glycoprotein (MOG), glial fibrillary acidic protein (GFAP) and p75 nerve growth factor receptor (p75NGFR) (left, all green; DAPI-stained nuclei, blue). Cells labelled with secondary antibody alone (2° Ab) were used as controls to determine unspecific staining. Scale bars, 10 μm. Expression in all cell lines was confirmed by western blotting (right). Lysate from HEK293T wild-type cells (HEK^{WT}) was used as control. All proteins except MBP were expressed in SW10 cells and its derivatives. For uncropped gels, see Supplementary Information File 2. **b**, Western blot (developed with POM2) of HEK 293T cells transfected with expression plasmids for wild-type murine PrP^C or for PrP^C bearing lysine-to-alanine substitutions in the KKRPK and QGSPG motifs (lanes 3 and 4, respectively). The mutations did not affect the biogenesis and processing of PrP^C **c**, Western blot of medium collected from the cells shown in **b**. FT fragments bearing the mutations were released into the medium similarly to wild-type FT. For uncropped gels, see Supplementary Information File 1. **d**, $SW10_{\Delta PrP}$ cells were treated with conditioned medium from HEK293T cells transfected with an empty vector (HEK^{empty}), with PrP^C (HEK^{PrP}), or with full-length PrP^C versions in which the QGSPG ($HEK^{PrP}_{\sum QGSPG}$) or KKRPK ($HEK^{PrP}_{\sum KKRPK}$) motifs were substituted ($\sum$) with alanines. The charge neutralization within the KKRPK motif abrogated the cAMP induction. **e**, $SW10_{\Delta PrP}$ cells were treated with $FT_{23–50}$ (2 μM) or a Col4-derived 21-meric synthetic peptide containing either the GPRGKPG domain or its alanine-substituted variant (AAAGAAG; both 8 μM). Both $FT_{23–50}$ and the native Col4 peptide, but not the alanine-substituted peptide (Ala-Col4), induced cAMP. Data are representative of three biologically independent experiments; statistical significance was evaluated by unpaired Student's *t*-test. Error bars show s.e.m.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 7 | *Prnp*$^{ZH3/ZH3}$ and *Gpr126*$^{\Delta Schwann}$ mice display comparable demyelination phenotypes. a, Transmission electron micrographs of sciatic nerves from 14-month-old *Prnp*$^{ZH1/ZH1}$ mice (ZH1). Black arrowhead, thinly myelinated axons; white arrowhead, abnormal cytoplasmic Schwann cell protrusions; boxes, loss of axon–Schwann cell interactions; asterisk, initial onion bulb formation. Scale bar, 2 μm in upper left panel; 500 nm in all other panels. **b**, **c**, Quantification of unmyelinated axons in Remak bundles was performed manually by counting the number of axons in the bundles from electron microscopy images (1,500× magnification, 10 images per mouse were analysed and three mice per genotype were used in total). The bundles were further sorted into three categories: <10 axons, 10–20 axons and >20axons per bundle. Comparisons were performed between either BL6 and *Prnp*$^{ZH3/ZH3}$ (**b**) or *Gpr126*$^{fl/fl}$ (WT) and *Dhh*^{Cre}::*Gpr126*$^{fl/fl}$ (*Gpr126*$^{\Delta Schwann}$) mice (**c**; all mice were 13 months old). Both *Prnp*$^{ZH3/ZH3}$ and *Gpr126*$^{\Delta Schwann}$ mice showed a similar inclination towards a decrease in the number of axons per bundle. Statistical significance was established by performing a two-way ANOVA with Bonferroni correction. **d**, Onion bulb-like structures were quantified from electron microscopy images (1,500× magnification, 10 images per mouse were analysed and three mice per genotype were used in total) of either BL6 and *Prnp*$^{ZH3/ZH3}$ or *Gpr126*$^{fl/fl}$ (WT) and *Dhh*^{Cre}::*Gpr126*$^{fl/fl}$ (*Gpr126*$^{\Delta Schwann}$) mice. These onion bulb-like structures were prevalent only in *Prnp*$^{ZH3/ZH3}$ and *Gpr126*$^{\Delta Schwann}$ mice, with *Gpr126*$^{\Delta Schwann}$ exhibiting more. Error bars show s.e.m.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 8 | FT is myelinotrophic in both zebrafish and mice. a, Immunofluorescence for Mbp (green) in the posterior lateral line nerve of wild-type zebrafish larvae. AcTub: acetylated tubulin (red) labelling axons. Scale bar, 20 μm. **b**, *gpr126*^{st49} hypomorphic mutant larvae were treated with vehicle (DMSO) or FT_{23-50} (20 μM) at 50–55 hpf and immunostained at 5 dpf for Mbp (green). AcTub: acetylated tubulin (red) labelling axons. Scale bar, 20 μm. FT_{23-50} did not alter Mbp immunofluorescence. **c**, FT_{23-50} or $FT^{K\rightarrow A}_{23-50}$ was intravenously administered to *Prnp*$^{ZH1/ZH1}$ and BL6 mice (600 μg per mouse, 20 min). After FT_{23-50} injection, cAMP levels in *Prnp*$^{ZH1/ZH1}$ mice increased to levels approaching those of BL6 mice. Each dot represents one mouse. **d–f**, cAMP also spiked in hearts of mice injected with FT_{23-50} but not $FT^{K\rightarrow A}_{23-50}$ (**d**). $^{*}P < 0.05$; $^{**}P < 0.01$. **c**, **d**, FT_{23-50} or $FT^{K\rightarrow A}_{23-50}$ was injected intravenously into 10–16-week-old *Prnp*$^{ZH1/ZH1}$ or BL6 mice (600 μg per animal, 20 min). cAMP levels in kidneys (**e**) and brain (**f**) showed no significant changes. Each dot represented an individual mouse; statistical significance was evaluated by unpaired Student's *t*-test. Error bars show s.e.m.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 1 | Sequences of synthetic peptides used in the present study. The collagen-4 homology domain necessary for cAMP induction is highlighted in yellow

Name	Sequence
FT_{23-50}	KKRPKPGGWNTGGSRYPGQGSPGGNRYP
FT_{23-50}-HA	KKRPKPGGWNTGGSRYPGQGSPGGNRYPYPYDVPDYA
FT_{34-69}-HA	PGQGSPGGNRYPPQGGTWGQPKGGGWGQYPYDVPDYA
FT_{51-94}-HA	PQGGTWGQPHGGGWGQPHGGSWGQPHGGSWGQPHGGGW
FT_{83-109}-HA	PHGGGWGQGGGTHNQWNKPSKPKTNLKYPYDVPDYA
FT_{94-110}-HA	GGGTHNQWNKPSKPKTNLKHYPYDVPDYA
FT_{94-110}	GGGTHNQWNKPSKPKTNLKH
AlaQGSPG	KKRPKPGGWNTGGSRYPGAAAAAGNRYP
AlaKKRPK	AAAPAPGGWNTGGSRYPGQGSPGGNRYP
ColIV	GPRGKPGVDGYNGSRGDPGYP
ColIV-Mut	AAAGAAGVDGYNGSRGDPGYP
FT_{23-34}	KKRPKPGGWNTG
SFT_{23-34}	AAAGAAGGWNTG

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

LETTER

doi:10.1038/nature19073

DWARF14 is a non-canonical hormone receptor for strigolactone

Ruifeng Yao[1]*, Zhenhua Ming[2]*, Liming Yan[2]*, Suhua Li[1]*, Fei Wang[1], Sui Ma[1], Caiting Yu[1], Mai Yang[1], Li Chen[1], Linhai Chen[3], Yuwen Li[1], Chun Yan[1], Di Miao[1], Zhongyuan Sun[1], Jianbin Yan[1], Yuna Sun[4], Lei Wang[5], Jinfang Chu[5], Shilong Fan[1], Wei He[6], Haiteng Deng[1], Fajun Nan[3], Jiayang Li[5], Zihe Rao[2,4], Zhiyong Lou[2] & Daoxin Xie[1]

Classical hormone receptors reversibly and non-covalently bind active hormone molecules, which are generated by biosynthetic enzymes, to trigger signal transduction. The α/β hydrolase DWARF14 (D14), which hydrolyses the plant branching hormone strigolactone and interacts with the F-box protein D3/MAX2, is probably involved in strigolactone detection[1–3]. However, the active form of strigolactone has yet to be identified and it is unclear which protein directly binds the active form of strigolactone, and in which manner, to act as the genuine strigolactone receptor. Here we report the crystal structure of the strigolactone-induced AtD14–D3–ASK1 complex, reveal that *Arabidopsis thaliana* (At)D14 undergoes an open-to-closed state transition to trigger strigolactone signalling, and demonstrate that strigolactone is hydrolysed into a covalently linked intermediate molecule (CLIM) to initiate a conformational change of AtD14 to facilitate interaction with D3. Notably, analyses of a highly branched *Arabidopsis* mutant *d14-5* show that the AtD14(G158E) mutant maintains enzyme activity to hydrolyse strigolactone, but fails to efficiently interact with D3/MAX2 and loses the ability to act as a receptor that triggers strigolactone signalling *in planta*. These findings uncover a mechanism underlying the allosteric activation of AtD14 by strigolactone hydrolysis into CLIM, and define AtD14 as a non-canonical hormone receptor with dual functions to generate and sense the active form of strigolactone.

The classical hormone receptors that have been reported so far reversibly bind their active hormone molecules, which are generated through sequential actions of biosynthetic enzymes[4]. In the plant kingdom, all known hormones non-covalently bind their receptors to initiate signalling, and dissociate from the receptors without alteration[5–10] (for further information, see Supplementary Information refs 43–61). Strigolactones (SL) are plant hormones that play a vital role in the control of plant branching[11–13], and also act as rhizospheric signals for communication with symbiotic fungi[14] and parasitic plants[15–17]. *MORE AXILLARY GROWTH2* (*MAX2*) and *DWARF14* (*AtD14*) in *Arabidopsis*[18–20], their orthologues in rice (*D3* and *OsD14*)[21,22] and other species[1,15–17,23], are essential signalling components to regulate the SL-repressed plant branching. *D14* encodes an α/β hydrolase that hydrolyses GR24 (a synthetic SL analogue) into an intermediate 2,4,4,-trihydroxy-3-methyl-3-butenal (TMB)[24], the final products hydroxymethyl butenolide (D-OH)[2,24] and tricyclic lactone (ABC-OH)[1,2,24], all of which lose their SL biological activity[1,25]. *MAX2*/*D3* encodes an F-box protein, which interacts with D14 in a SL-dependent manner[1,3,25–27] and recruits various repressors (such as rice D53 and *Arabidopsis* SMXL6/7/8) for 26S proteasome-mediated degradation[25–27]. The D14–D3/MAX2 complex[28], D14–D53/SMXLs complex[29], D3/MAX2 protein[26], particularly, D14 protein[1,2,15–17,19] were previously proposed to be involved in SL perception. However, the active form of SL has not been determined, and it remains unknown which protein directly binds the active form of SL, and in which manner, to act as the SL receptor that triggers SL signalling.

To investigate the SL perception mechanism, we purified AtD14 and D3, which share high homology with MAX2 (Extended Data Fig. 1a) and rescue the *max2* mutant (Extended Data Fig. 1b, c), assembled the GR24-induced complex of AtD14 and D3 (co-expressed with ASK1 protein[5,8]), and solved the crystal structure of this complex (Fig. 1a, b and Extended Data Table 1). This structure presents a seahorse-shaped organization, in which AtD14 acts as the head part, the leucine-rich repeat domain of D3 (D3 LRR) forms the body, and the F-box motif of D3 (D3 F-box) together with ASK1 serve as the tail. AtD14 is attached to the C terminus of D3 LRR with its catalytic cavity facing to D3 LRR. Electron density maps (Extended Data Fig. 2a) suggest that a small molecule is sealed inside the closed catalytic cavity of AtD14 without contacting D3 (Fig. 1a): this small molecule fixes its two arms through stable contacts with the $N^{\varepsilon 2}$ atom of H247 and the hydroxyl oxygen of S97 of AtD14. Consistent with these structural observations, GR24 hydrolysis assays showed that the presence of D3 inhibited the AtD14-mediated GR24 hydrolysis to attenuate the release of D-OH (Extended Data Fig. 2b), suggesting that D3 binds AtD14 to retain a hydrolytic D-ring-derived intermediate inside the closed catalytic cavity of AtD14.

To further identify the small molecule sealed inside AtD14, we collected the GR24-induced AtD14–D3–ASK1 complex using size-exclusion chromatography (Fig. 1b) to separate AtD14 for tandem mass spectrometry (MS/MS) analysis. Peptide matching from MS/MS spectra identified a modified peptide of AtD14 with a molecular weight corresponding to the $C_5H_5O_2$ modification covalently linked to H247 (Extended Data Fig. 3b), whereas such a modification was not detected on the hydrolytic mutants AtD14(S97A) or AtD14(H247A) (Extended Data Fig. 4c–e). All of the reported natural SLs contain the same enol ether bridge-linked D-ring as GR24 but have variable groups to replace the ABC ring[3,12,25]. The biologically active compounds, 5-deoxystrigol (5DS, a native SL) and 4-Br debranone (4BD, an SL analogue) (Extended Data Fig. 3a), both effectively induced AtD14–D3 complex formation (Extended Data Fig. 3c) and generated the same $C_5H_5O_2$ modification as GR24 on AtD14 (Extended Data Fig. 3d, e). In contrast, the biologically inactive compounds[17,25] carba-GR24 and D-OH, without the ability to induce the AtD14–D3/MAX2 interaction, failed to generate the $C_5H_5O_2$ modification on AtD14 (Extended Data Fig. 5). These results suggest that various biologically active SL molecules are hydrolysed into the same D-ring-derived intermediate sealed inside the D3-bound AtD14. In support of this, further MS/MS analysis using 4BD with the $^2H_3$-labelled D-ring ($^2H_3$-4BD) identified a $^2H_3$-labelled $C_5H_5O_2$ modification on AtD14 (Fig. 1c, d and Extended Data

[1]Tsinghua-Peking Joint Center for Life Sciences, and MOE Key Laboratory of Bioinformatics, School of Life Sciences, Tsinghua University, Beijing 100084, China. [2]School of Medicine and Collaborative Innovation Center of Biotherapy, Tsinghua University, Beijing 100084, China. [3]Shanghai Institute of Materia Medica, Chinese Academy of Science, Shanghai 201203, China. [4]Institute of Biophysics, Chinese Academy of Science, Beijing 100101, China. [5]State Key Laboratory of Plant Genomics, and National Center for Plant Gene Research (Beijing), Institute of Genetics and Developmental Biology, Chinese Academy of Science, Beijing 100101, China. [6]School of Pharmaceutical Sciences, Tsinghua University, Beijing 100084, China.
*These authors contributed equally to this work.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

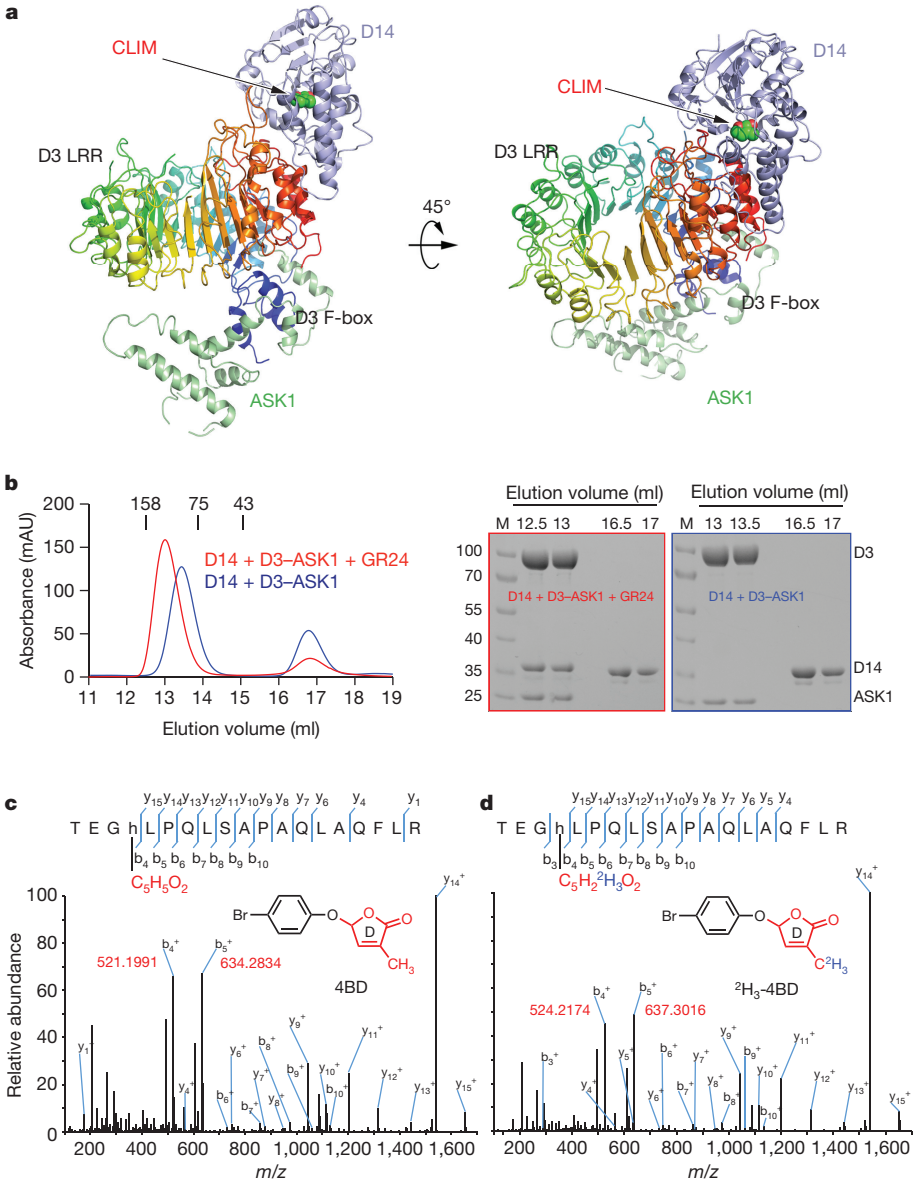


Figure 1 | Overall structure of AtD14–D3–ASK1 in complex with an active SL intermediate. a, Overall structure of AtD14–D3–ASK1 in complex with a GR24 hydrolysis intermediate CLIM. AtD14, ASK1 and D3 are coloured as light blue, pale green and rainbow scheme (N to C terminus from blue to red), respectively. CLIM is shown as space-fill representation. **b**, Size-exclusion chromatography analysis (left panel) of the interaction between AtD14 and D3–ASK1; the elution volumes of the molecular weight markers are indicated on the top. SDS–PAGE analysis (right panel) of peak fractions from the left panel; M, molecular weight ruler (kDa). **c**, **d**, MS/MS spectra of a bivalently charged peptide (244-TEGhLPQLSAPAQLAQFLR-262) of AtD14 at mass-to-charge ratio (*m*/*z*) 1087.0736 or 1088.5828 corresponding to the mass of $C_5H_5O_2$ or $C_5H_2{}^2H_3O_2$ modification of the peptide, which was isolated from AtD14 in the 4BD- (**c**) or $^2H_3$-labelled 4BD-induced (**d**) AtD14–D3 complex with in-solution digestion method, respectively. Labelled peaks correspond to masses of y and b ions of the modified peptide. Lowercase 'h' indicates the modified H247.

Fig. 6). Moreover, the same $C_5H_5O_2$ modification was also detected *in planta* when plants were treated with GR24 (Extended Data Fig. 4a) or 5DS (Extended Data Fig. 4b).

Together with the previously determined SL hydrolysis compounds[1–3,24] and the general principle of Ser–His–Asp catalytic triad, a combination of above-mentioned findings enable us to rationally deduce the chemical structure of this D-ring-derived small molecule CLIM in our AtD14–D3–ASK1 crystal (Extended Data Fig. 2c, d and Supplementary Notes), suggesting that CLIM is generated by AtD14 during SL hydrolysis and covalently sealed inside AtD14 to act as the active SL molecule that initiates SL signalling.

Notably, the architecture of the binding pocket for CLIM in our crystal markedly differs from that of GR24, TMB and D-OH in the previously reported D3-unbound OsD14 crystals, which possess the similar overall structure as apo AtD14/OsD14 protein[1–3,24]. Compared to the apo AtD14 (open state), AtD14 in the GR24-induced AtD14–D3 complex (closed state) has experienced significant conformational changes, referred to as the open-to-closed transition (Fig. 2). The open state contains an unclosed lid with four top helices (αT1–αT4), and a large open pocket (420 Å^3) (Fig. 2a–d and Extended Data Fig. 7a–f) compatible with binding large molecules (such as GR24). However, the closed state contains a collapsed lid with only three helices (αT1, αT3 and αT4), and a small closed pocket (80 Å^3) (Fig. 2e–h) suitable for embedding small molecules (such as CLIM). During the open-to-closed state transition, αT2 rotates towards αT1 while αT3 rotates towards αT4, coupled by the helix-to-coil transition of αT2 and the coil-to-helix transition of the loop between αT1 and αT2, making the pocket smaller and closed to generate the flat tri-helix lid compatible for binding D3 (Fig. 2e–g). Notably, S97 and H247 stay almost in the original positions of the catalytic triad (S97–H247–D218),

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

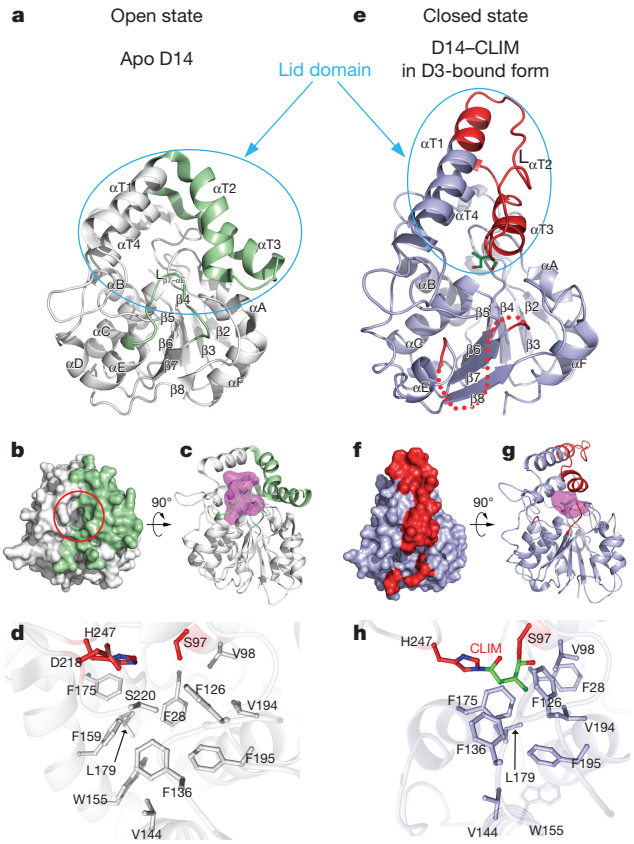


Figure 2 | Allosteric activation of AtD14 for the open-to-closed conformational transition. a–h, Motif labelling and structural comparison between apo AtD14 (**a**, PDB code: 4IH4) and AtD14–CLIM in D3-bound form (**e**), focusing on the shape (**b** and **f**), volume (**c** and **g**) and details (**d** and **h**) of the catalytic cavity. **a–d** Apo AtD14 in the open state; **f–h**, AtD14–CLIM in D3-bound form in the closed state.

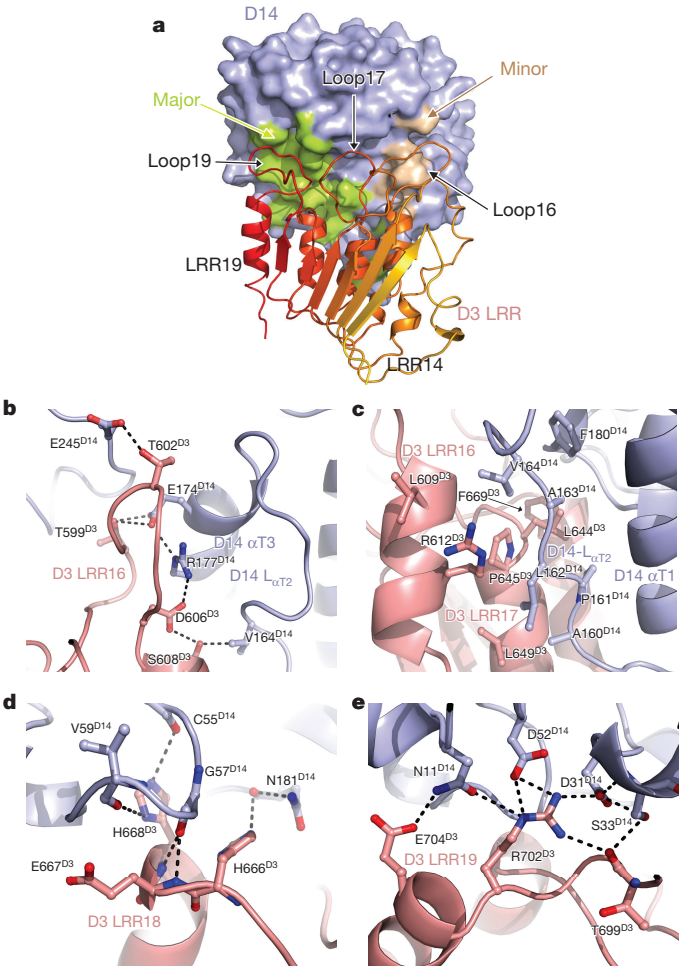


Figure 3 | Structural analyses of the AtD14-D3 interaction. **a**, An overview of the AtD14–D3 interface. Loop16, loop17 and loop19, which are involved in the AtD14–D3 interaction, are labelled. The minor and major parts of the interface in AtD14 are highlighted by sand and green, respectively. **b–e**, Structural interactions between AtD14 and D3. The structural elements of AtD14 and D3 are coloured light blue and salmon, respectively. Residues are labelled with a superscript ^{AtD14} or ^{D3} to indicate which protein they originate from. Dashed lines indicate hydrogen-bonding interactions, except for van der Waals forces between $S608^{D3}$ and $V164^{AtD14}$.

and constitute a hydrophobic cavity, together with a subset of residues (including F28, F126, F175, L179 and V194) (Fig. 2h), to seal CLIM inside AtD14. However, the $L_{\beta7-\alpha E}$ loop (the loop region between β7 and αE) containing D218 of the catalytic triad (Fig. 2a) moves away and becomes disordered in the D3-bound AtD14 (Fig. 2e), which would be expected to disrupt the S97–H247–D218 catalytic triad of AtD14 to prevent CLIM from further hydrolysis into final product D-OH in the D3-bound AtD14. It is possible that this destabilized $L_{\beta7-\alpha E}$ in the D3-bound AtD14 is involved in interaction with other proteins such as the repressors D53/SMXLs that interact with AtD14/OsD14 in a SL-dependent manner[1,3,25–27].

We further solved the crystal structure of AtD14-unbound D3–ASK1 to make a structural comparison with the AtD14-bound D3–ASK1. Both structures contain the N-terminal tri-helical F-box motif bound to ASK1, and the C-terminal LRR domain (D3 LRR) with 19 LRRs packing in tandem and assembling into a solenoid with an overall horseshoe-shaped structure (Extended Data Fig. 8a–d), which is similar to TIR1–ASK1 and COI1–ASK1 complexes[5,8]. As a notable structural feature of AtD14-bound D3, LRR15, LRR16, LRR17 and LRR19 each have an unusual long loop between their α-helix and β-strand (Extended Data Fig. 8a, b). Loop16, loop17 and loop19 are involved in the interaction with AtD14 (Fig. 3a) and in turn stabilized by AtD14, whereas loop15 is not directly involved in contacting AtD14 although stabilized by AtD14 (Extended Data Fig. 8c, d). In contrast, most of LRR13 is disordered in both structures, implying that LRR13 (and loop15) are possibly involved in targeting another protein.

Further structural analyses reveal a well-defined interface to mediate the AtD14–D3 interaction. The C-terminal LRRs of D3 extensively bind AtD14 in the closed state (on the lid domain as well as some short helices and loops near the lid) to form an AtD14–D3 interface with an area of 3,013 Å^2 (Fig. 3a). This interface has two separate parts (termed the minor interface and major interface, on the basis of their sizes). The minor interface contains an extensive intermolecular hydrogen bond network, involving four critical residues ($T599^{D3}$ and $D606^{D3}$ of loop 16, $E174^{AtD14}$ and $R177^{AtD14}$ of αT3) (Fig. 3b). This network is further stabilized by $T602^{D3}$–$E245^{AtD14}$ hydrogen bonding and $S608^{D3}$–$V164^{AtD14}$ van der Waals contact. The major interface is further divided into two distinct areas: (1) a hydrophobic area (Fig. 3c), with predominant van der Waals interactions, consists of residues from D3–LRR ($L609^{D3}$, $R612^{D3}$, $L644^{D3}$, $P645^{D3}$, $L649^{D3}$ and $F669^{D3}$) and the AtD14 lid domain ($A160^{AtD14}$ to $V164^{AtD14}$ and $F180^{AtD14}$); (2) a hydrophilic area (Fig. 3d, e), with many hydrogen-bond interactions, consists of residues from two helices of D3 LRR (LRR18 and LRR19) and a AtD14 surface region that is covered by αT3 in the AtD14 open state but exposed in the closed state. D3 LRR18 contributes to AtD14–D3 interaction through a tandem hydrogen bonding zipper ($N181^{AtD14}$–$H666^{D3}$–$G57^{AtD14}$–$E667^{D3}$–$H668^{D3}$–$C55^{AtD14}$/$V59^{AtD14}$) (Fig. 3d), whereas D3 LRR19 contributes to AtD14–D3 interaction through a cascade of intermolecular hydrogen bonds ($E704^{D3}$–$N11^{AtD14}$–$R702^{D3}$–$D52^{AtD14}$–$R702^{D3}$–$D31^{AtD14}$–$S33^{AtD14}$–$T699^{D3}$–$R702^{D3}$)

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

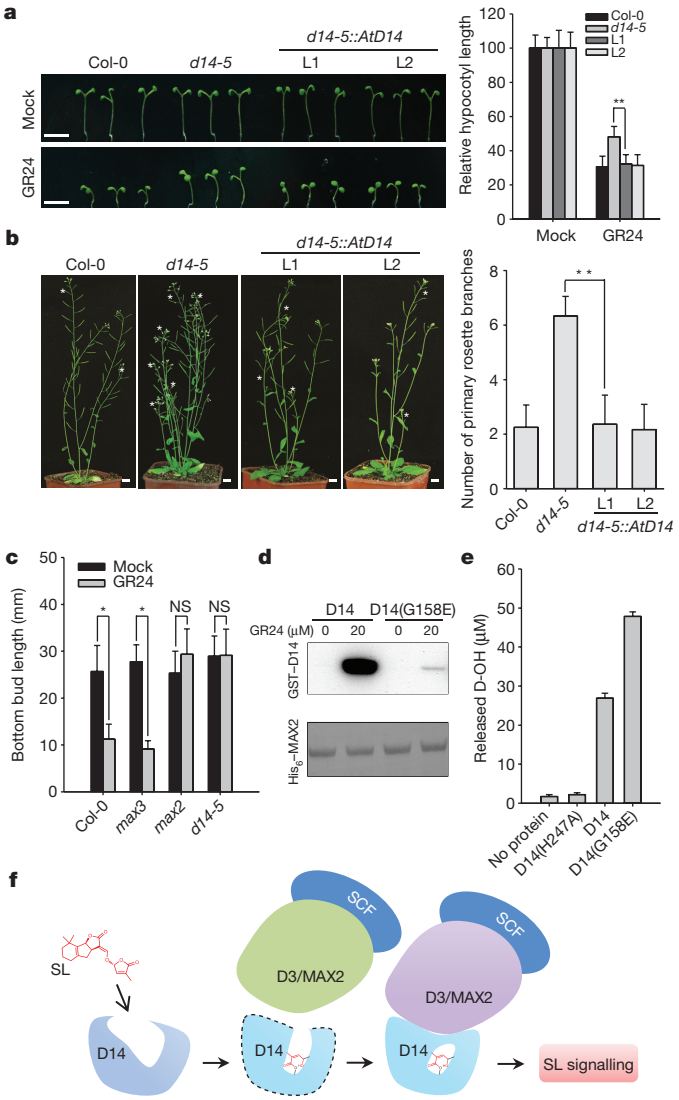


Figure 4 | AtD14(G158E) hydrolyses SL but fails to efficiently bind MAX2 for triggering SL signalling. a, **b**, The *d14-5* mutant exhibits less sensitivity to GR24 inhibitory hypocotyl elongation (**a**) and displays significantly more primary branches (**b**). **a**, Representative hypocotyl phenotype (left panel) of Col-0, *d14-5* and two independent transgenic lines *d14-5::AtD14* line 1(L1) and *d14-5::AtD14* L2; and relative hypocotyl lengths (right panel) of the indicated seedlings. Scale bars, 5 mm. The images are representative of 20 plants for each genotype, each treatment and each biological replicate. **b**, Representative branching phenotype (left panel) and quantitative analysis of primary rosette branches (right panel) of the indicated plants. Scale bars, 1 cm. The images are representative of 20 plants for each genotype and each biological replicate. Data in both **a** and **b** are means ± s.d. ($n = 3$ biological replicates of 20 plants per genotype); $^{**}P < 0.01$ (ANOVA). **c**, The branch elongation of the bottom bud of the excised stem segments is inhibited by GR24 in Col-0 and *max3-11*, but not in *max2-3* and *d14-5*. Data are means ± s.d. ($n = 3$ biological replicates of 12 segments per genotype); NS, not significant; $^{*}P < 0.05$ (ANOVA). **d**, AtD14(G158E) fails to efficiently bind MAX2. Pull-down assays using recombinant His_6–MAX2 and GST–AtD14 or GST–AtD14(G158E) in the absence or presence of GR24. Full blots are shown in Supplementary Fig. 1. **e**, AtD14(G158E) exhibits higher activity to hydrolyse (+)-GR24 into D-OH. Released D-OH in each indicated reaction mixture were quantitated by LC-MS/MS. Data are means ± s.d. ($n = 6$). **f**, A simplified model of SL perception. AtD14 in its open state docks SL in the catalytic cavity, hydrolyses SL into a D-ring-derived molecule (CLIM), which is covalently sealed inside the catalytic centre of AtD14. AtD14 undergoes conformational change suitable for interaction with D3/MAX2-based SCF complex, which may in turn strengthen this conformational change, to trigger SL signal transduction.

(Fig. 3e). Key residues in both minor and major interfaces are largely conserved (Extended Data Figs 1a and 7g), and the amino acid substitution of the critical residues (such as F180 (ref. 3), P161, E174 and R177 of AtD14; and D606, L644 and R702 of D3) severely attenuates GR24-induced AtD14–D3 interaction (Extended Data Fig. 8e, f). Taken together, our structural analyses reveal a well-defined interface that mediates the AtD14–D3 interaction.

Genetic analysis of an *Arabidopsis* mutant allele *d14-5*, with the Gly158Glu substitution in AtD14 (AtD14(G158E)), showed that *d14-5* exhibited reduced sensitivity in GR24-inhibitory hypocotyl elongation (Fig. 4a) and axillary bud growth (Fig. 4c), and displayed increased primary branches (Fig. 4b), suggesting that SL signalling is severely impaired in *d14-5*. Structural analyses suggest that the G158 residue participates in the construction of a π-turn structure (A154–W155–V156–H157–G158–F159) (Extended Data Fig. 9d, e). This π-turn structure, with an intrinsic capacity to stop further extension of the αT1 C terminus during the open-to-closed transition of AtD14 (Fig. 2a, e), contributes to stabilizing proper conformation of the AtD14 lid that is essential for binding D3/MAX2 (Fig. 4f and Extended Data Fig. 9e). G158E substitution would be expected to impair the closed state of the AtD14 lid to attenuate interaction between AtD14(G158E) and D3/MAX2. Indeed, AtD14(G158E) failed to efficiently interact with MAX2/D3 in the presence of GR24 (Fig. 4d and Extended Data Fig. 9b), suggesting that G158E substitution attenuates the GR24-induced interaction between AtD14 and D3/MAX2. Notably, AtD14(G158E) efficiently interacted with the signalling repressor SMXL6 in the presence of GR24 (Extended Data Fig. 9c), suggesting that AtD14(G158E) is able to undergo some conformational changes for binding SMXL6 in the presence of GR24. Notably, the SL hydrolysis assays showed that AtD14(G158E) maintains intact activity to mediate GR24 hydrolysis. In addition, AtD14(G158E) hydrolysed GR24 faster than AtD14 (Fig. 4e), which is consistent with our structural analyses that G158E substitution would impair the closed state of the AtD14(G158E) lid to enable more efficient GR24 uptake and D-OH release. These results demonstrate that the G158E substitution in AtD14(G158E) attenuates perception function to impair SL signalling, but maintain the ability to mediate SL hydrolysis, providing an independent line of evidence that AtD14 has dual functions as an enzyme that hydrolyses SL and a receptor that initiates SL signalling.

Previous crystallographic studies found that, similar to apo D14, all of the D3-unbound D14 in complex with GR24, TMB or D-OH are in the open state without overall conformational shift[1–3,24,30]. Our work has uncovered a closed state of AtD14 in the SL-induced AtD14–D3 complex. This closed state is probably attributable to D3 binding: AtD14 hydrolyses SLs into the D-ring-derived intermediate CLIM that covalently binds inside the catalytic pocket of AtD14, and undergoes conformational transition for interaction with D3, which may in turn strengthen this conformational change, to trigger SL signalling (Fig. 4f). Distinct from all known active phytohormone molecules that are first generated by biosynthesis enzymes and then non-covalently and reversibly bound by their receptors[5–10], CLIM represents a novel active hormone molecule that is generated by AtD14-mediated SL hydrolysis and sealed inside AtD14 through covalent links with H247 and probably S97, although direct evidence is needed to further validate the S97 linkage. Previous studies[25–27] have not yet elucidated the sequence of interactions among D14, D3/MAX2 and D53/SMXLs. Further structural and functional investigations on the D14–D3/MAX2–D53/SMXLs complex assembly will provide deeper mechanistic insights into the SL signalling pathway.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 29 January; accepted 1 July 2016.
Published online 1 August 2016.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

1. Hamiaux, C. *et al.* DAD2 is an α/β hydrolase likely to be involved in the perception of the plant branching hormone, strigolactone. *Curr. Biol.* **22,** 2032–2036 (2012).
2. Nakamura, H. *et al.* Molecular mechanism of strigolactone perception by DWARF14. *Nat. Commun.* **4,** 2613 (2013).
3. Zhao, L.-H. *et al.* Destabilization of strigolactone receptor DWARF14 by binding of ligand and E3-ligase signaling effector DWARF3. *Cell Res.* **25,** 1219–1236 (2015).
4. Norman, A. W. & Henry, H. L. *Hormones.* (Elsevier Science, 2014).
5. Tan, X. *et al.* Mechanism of auxin perception by the TIR1 ubiquitin ligase. *Nature* **446,** 640–645 (2007).
6. Murase, K., Hirano, Y., Sun, T. P. & Hakoshima, T. Gibberellin-induced DELLA recognition by the gibberellin receptor GID1. *Nature* **456,** 459–463 (2008).
7. Santiago, J. *et al.* The abscisic acid receptor PYR1 in complex with abscisic acid. *Nature* **462,** 665–668 (2009).
8. Sheard, L. B. *et al.* Jasmonate perception by inositol-phosphate-potentiated COI1–JAZ co-receptor. *Nature* **468,** 400–405 (2010).
9. Hothorn, M. *et al.* Structural basis of steroid hormone perception by the receptor kinase BRI1. *Nature* **474,** 467–471 (2011).
10. Hothorn, M., Dabi, T. & Chory, J. Structural basis for cytokinin recognition by *Arabidopsis thaliana* histidine kinase 4. *Nat. Chem. Biol.* **7,** 766–768 (2011).
11. Gomez-Roldan, V. *et al.* Strigolactone inhibition of shoot branching. *Nature* **455,** 189–194 (2008).
12. Umehara, M. *et al.* Inhibition of shoot branching by new terpenoid plant hormones. *Nature* **455,** 195–200 (2008).
13. Kretzschmar, T. *et al.* A petunia ABC protein controls strigolactone-dependent symbiotic signalling and branching. *Nature* **483,** 341–344 (2012).
14. Akiyama, K., Matsuzaki, K. & Hayashi, H. Plant sesquiterpenes induce hyphal branching in arbuscular mycorrhizal fungi. *Nature* **435,** 824–827 (2005).
15. Conn, C. E. *et al.* Plant evolution. Convergent evolution of strigolactone perception enabled host detection in parasitic plants. *Science* **349,** 540–543 (2015).
16. Toh, S. *et al.* Structure-function analysis identifies highly sensitive strigolactone receptors in *Striga*. *Science* **350,** 203–207 (2015).
17. Tsuchiya, Y. *et al.* Parasitic plants. Probing strigolactone receptors in *Striga hermonthica* with fluorescence. *Science* **349,** 864–868 (2015).
18. Stirnberg, P., van De Sande, K. & Leyser, H. M. MAX1 and MAX2 control shoot lateral branching in *Arabidopsis*. *Development* **129,** 1131–1141 (2002).
19. Waters, M. T. *et al.* Specialisation within the DWARF14 protein family confers distinct responses to karrikins and strigolactones in *Arabidopsis*. *Development* **139,** 1285–1295 (2012).
20. Chevalier, F. *et al.* Strigolactone promotes degradation of DWARF14, an α/β hydrolase essential for strigolactone signaling in Arabidopsis. *Plant Cell* **26,** 1134–1150 (2014).
21. Arite, T. *et al. d14*, a strigolactone-insensitive mutant of rice, shows an accelerated outgrowth of tillers. *Plant Cell Physiol.* **50,** 1416–1424 (2009).
22. Ishikawa, S. *et al.* Suppression of tiller bud activity in tillering dwarf mutants of rice. *Plant Cell Physiol.* **46,** 79–86 (2005).
23. Johnson, X. *et al.* Branching genes are conserved across species. Genes controlling a novel signal in pea are coregulated by other long-distance signals. *Plant Physiol.* **142,** 1014–1026 (2006).
24. Zhao, L. H. *et al.* Crystal structures of two phytohormone signal-transducing α/β hydrolases: karrikin-signaling KAI2 and strigolactone-signaling DWARF14. *Cell Res.* **23,** 436–439 (2013).
25. Jiang, L. *et al.* DWARF 53 acts as a repressor of strigolactone signalling in rice. *Nature* **504,** 401–405 (2013).
26. Zhou, F. *et al.* D14–SCF^{D3}-dependent degradation of D53 regulates strigolactone signalling. *Nature* **504,** 406–410 (2013).
27. Wang, L. *et al.* Strigolactone signaling in *Arabidopsis* regulates shoot development by targeting D53-Like SMXL repressor proteins for ubiquitination and degradation. *Plant Cell* **27,** 3128–3142 (2015).
28. Nelson, D. C. *et al.* F-box protein MAX2 has dual roles in karrikin and strigolactone signaling in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA* **108,** 8897–8902 (2011).
29. Umehara, M. *et al.* Structural requirements of strigolactones for shoot branching inhibition in rice and *Arabidopsis*. *Plant Cell Physiol.* **56,** 1059–1072 (2015).
30. Kagiyama, M. *et al.* Structures of D14 and D14L in the strigolactone and karrikin signaling pathways. *Genes Cells* **18,** 147–160 (2013).

Supplementary Information is available in the online version of the paper.

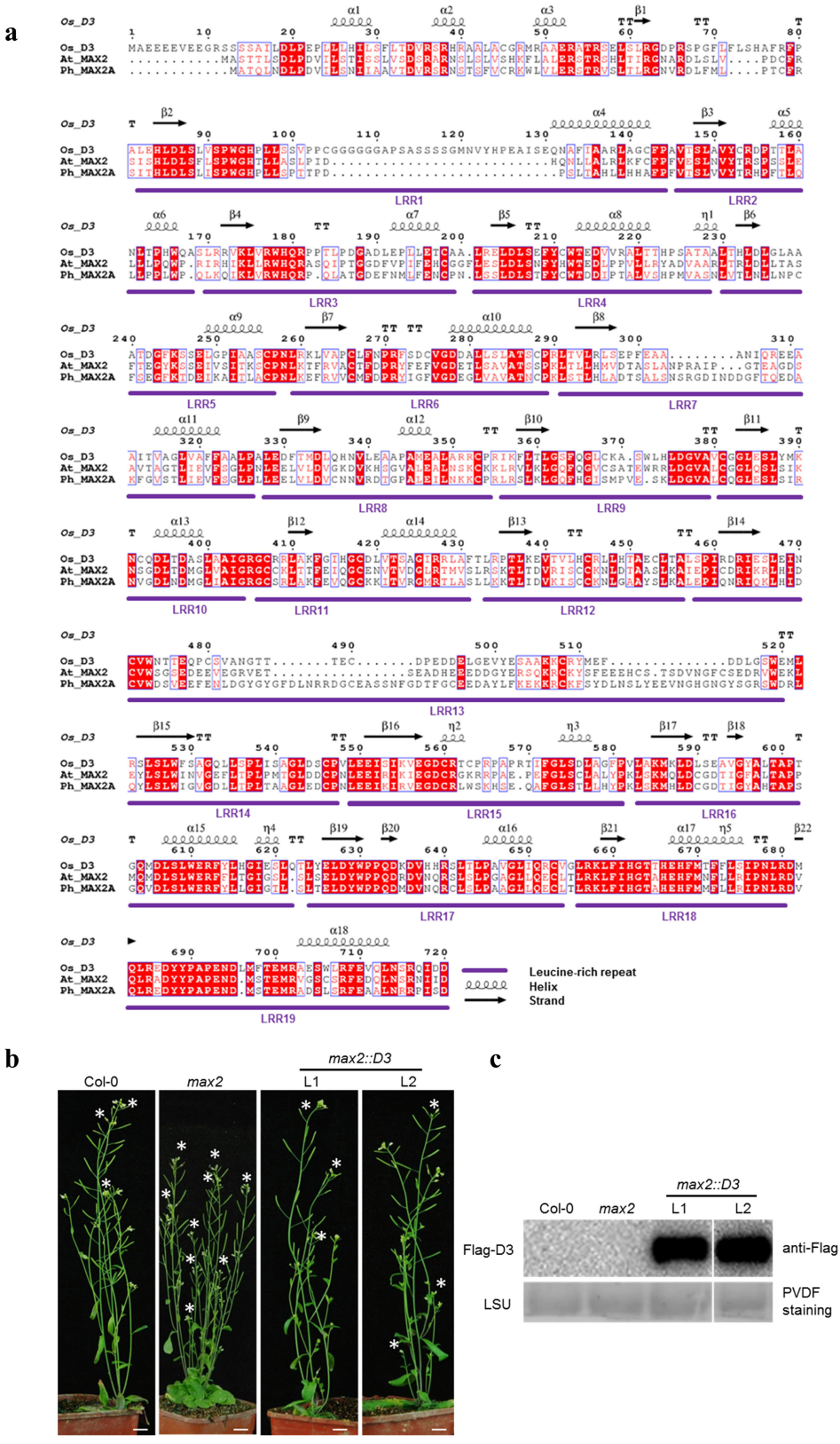
Acknowledgements We thank the staff (J. He, Q. Wang and F. Yu) at the Shanghai Synchrotron Radiation Facility (SSRF) beamline BL17U and the staff at Tsukuba Photon Factory beamline NE3A for help with data collection. We also thank G. Liu (School of Pharmaceutical Sciences), Z. Yin and H. Fu (Department of Chemistry) at Tsinghua University, B. Xia and P. Ding (School of Life Sciences), T. Luo (College of Chemistry and Molecular Engineering) at Peking University for valuable suggestions on chemistry analysis; Y. Ding, W. Wang and Y. Tian (Drug Facility, School of Pharmaceutical Sciences) at Tsinghua University for help with D-OH quantitation. This work was supported by the National Natural Science Foundation of China (Grant no. 91417302, 81322023, 31421001 and 91335204), the Ministry of Science and Technology (grant number 2013CB911100 and 2016YFA0500500), and the Ministry of Agriculture (grant number 2014ZX08011006).

Author Contributions D.X., Z.L. and Z.R. conceived the study. R.Y., and Z.M. performed molecular cloning, protein expression and purification, biochemical and structural experiments with assistance from L.Y., F.W., S.M., M.Y., C. Yan, S.L., J.Y., Li C., D.M., Z.S. and L.W.; Z.L. determined the crystal structure with assistance from L.Y., Z.M., S.F. and Y.S.; S.L. and R.Y. performed genetic assays with the help from C. Yu, Y.L.; Linhai C. and F.N. synthesized $^{2}H_{3}$-4BD; D.X., Z.L., Z.R., R.Y., Z.M., L.Y., S.L., J.L., H.D., W.H. and J.C. analysed data with help from all authors; D.X., Z.L., R.Y. and Z.M. wrote the paper with the inputs from all authors.

Author Information Crystallographic coordinates have been deposited in the Protein Data Bank under accession numbers 5HZG (CLIM–AtD14–D3–ASK1) and 5HYW (D3–ASK1). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.X. (daoxinlab@mail.tsinghua.edu.cn), Z.L. (louzy@mail.tsinghua.edu.cn) or Z.R. (raozh@mail.tsinghua.edu.cn).

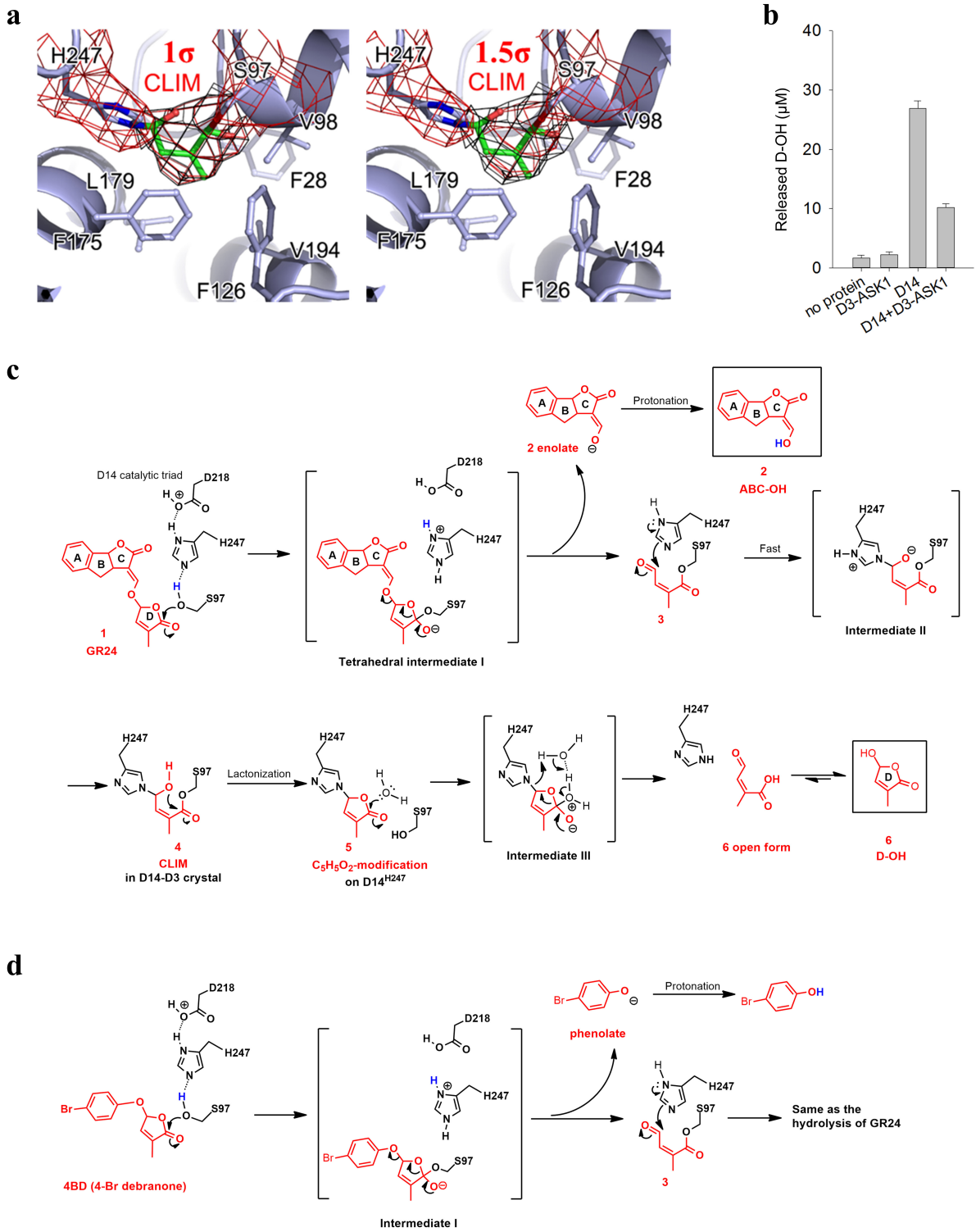
Reviewer Information *Nature* thanks P. McCourt, K. Snowden and the other anonymous reviewer(s) for their contribution to the peer review of this work.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 1 | D3 orthologues among different plant species and complementation of *Arabidopsis max2* by rice *D3*. **a**, Secondary structure elements of D3 in the crystal structure of D3–AtD14 complex are shown on top of *Oryza sativa* D3 sequence. Identical and conserved residues are highlighted by red and white grounds, respectively. The 19 LRRs are indicated by solid magenta rectangles. The GenBank accession numbers from top to bottom: *Oryza sativa* D3 (BAD69288), *Arabidopsis thaliana* MAX2 (NP565979) and *Petunia hybrida* MAX2 A (AEB97384). **b**, Branching phenotypes of seven-week-old Col-0, *max2-3* and two independent lines (L1 and L2) of *max2-3* plants expressing Flag-tagged rice D3 (*max2::D3*). Scale bar, 1 cm. The images are representative of 20 plants for each genotype. **c**, The Flag–D3 protein in corresponding plants (**b**) was detected by anti-Flag antibody (upper panel). Staining of Rubisco large subunit (LSU) served as a loading control (lower panel). Full blots are shown in Supplementary Fig. 1.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

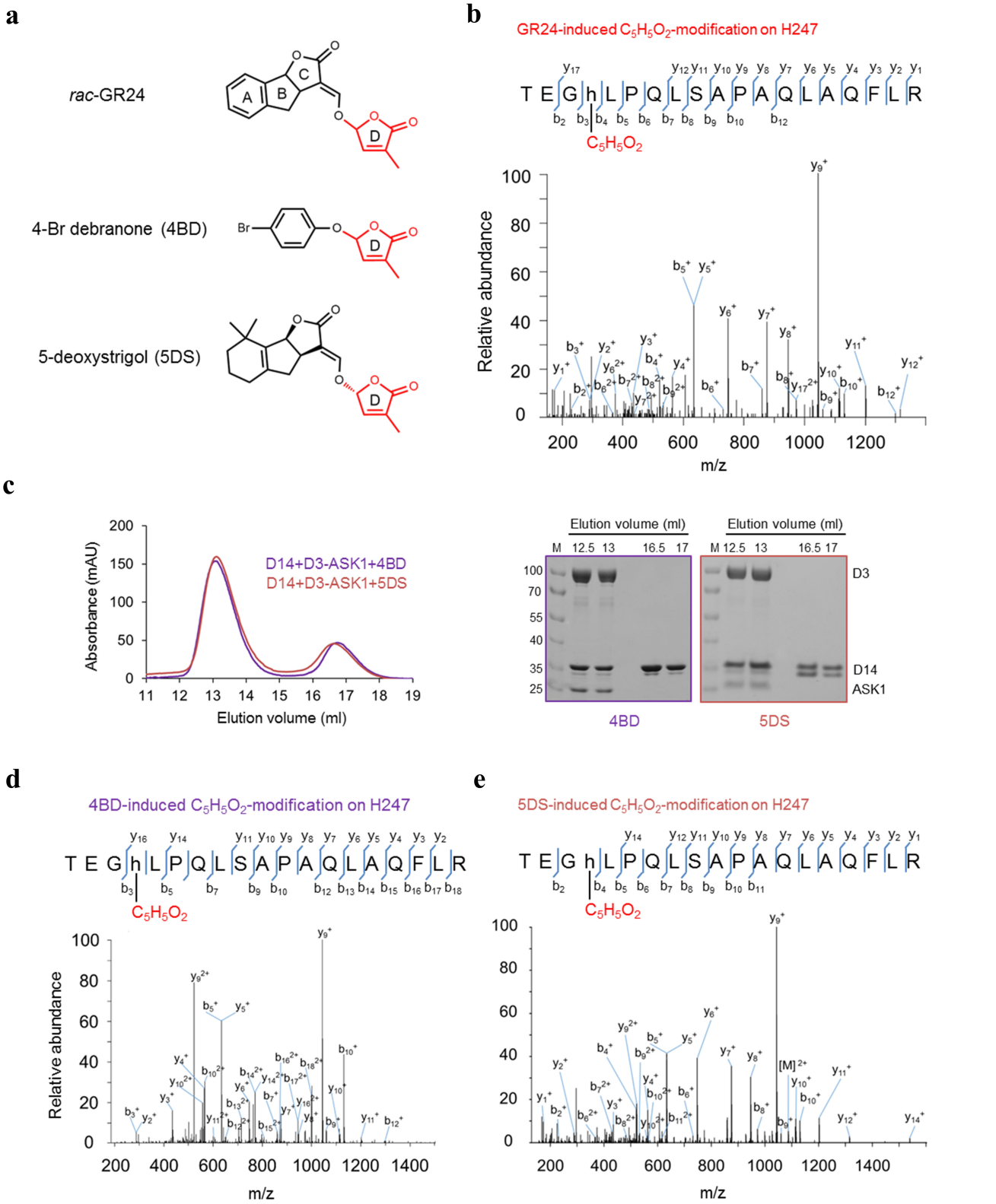


Extended Data Figure 2 | See next page for caption.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

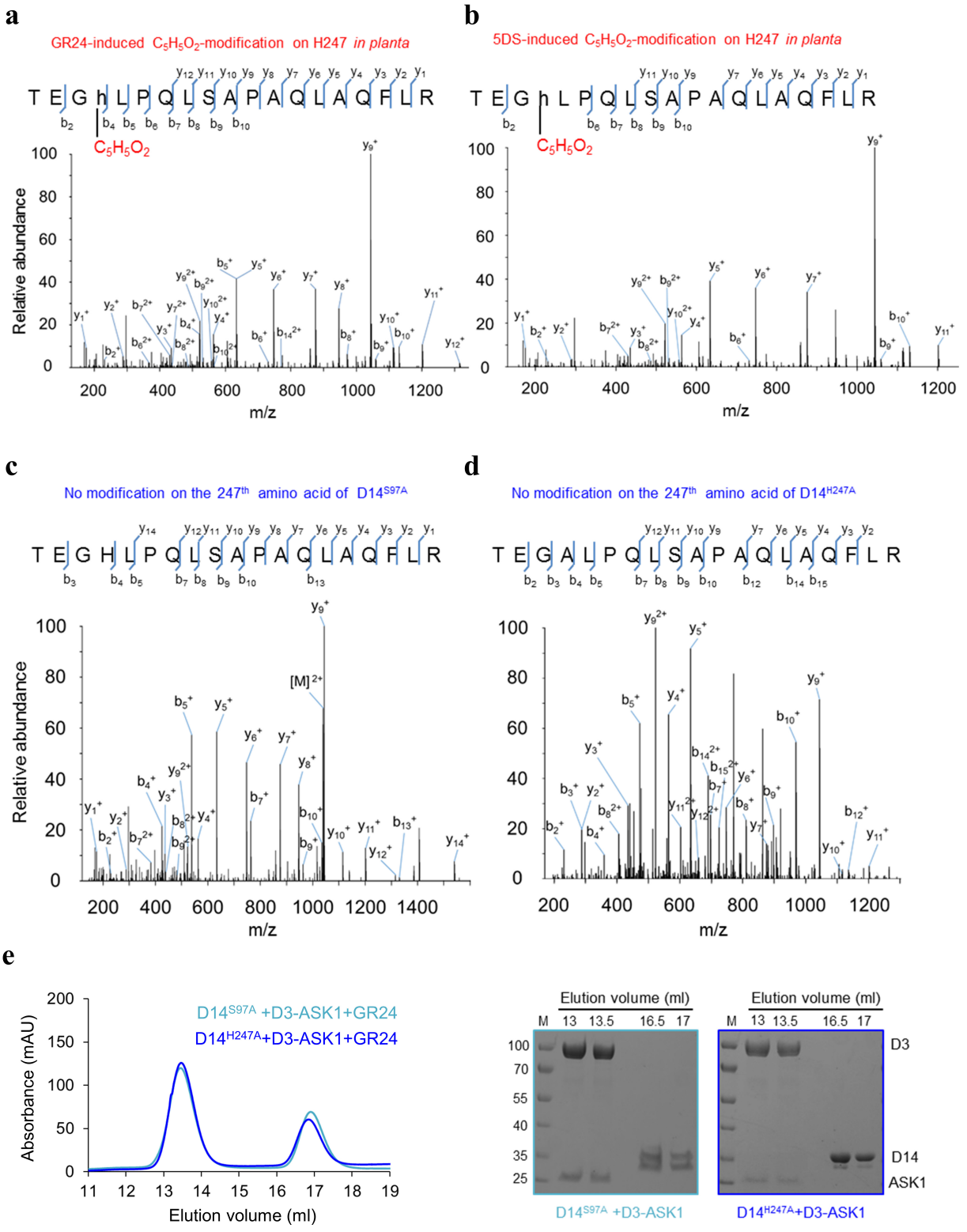
Extended Data Figure 2 | A proposed mechanism of AtD14-mediated SL hydrolysis. a, CLIM forms stable contacts with S97 and H247 in the catalytic centre of AtD14. CLIM, the nearby contacting residues and two catalytic residues (S97 and H247) are represented as sticks. The $2F_o - F_c$ omit map (black mesh) for CLIM is contoured at 1σ and the $2F_o - F_c$ electron density maps (red mesh) for CLIM, S97 and H247 are contoured at 1σ (left panel) and 1.5σ (right panel), respectively. **b**, D3 inhibits the release of D-OH during the AtD14-mediated GR24 hydrolysis. Released D-OH in each indicated reaction mixture after 60 min incubation at 37 °C were quantitated by LC-MS/MS. 50 μM GR24, 8 μM AtD14, and/or 15 μM D3 were used. Data are means ± s.d. ($n = 6$). **c**, Proposed mechanism of AtD14-mediated GR24 (compound **1**) hydrolysis. The reaction begins with the well-accepted nucleophilic attack by S97 in the catalytic triad[1,2,24,30]. Decomposition of the tetrahedral intermediate I resulted in the enolate form of ABC-OH (**2**) (which upon protonation gives compound **2**) and the S97-bound moiety **3**. The alternative direct hydrolysis of GR24 (**1**) is not supported by the fact that AtD14 S97 mutant is catalytically inactive[1]. Likewise, direct hydrolysis of the ester bond in S97-bound moiety **3** is an unlikely event. As the electron density shape (Extended Data Fig. 2a) and the $C_5H_5O_2$ modification (Fig. 1c, d and Extended Data Figs 3 and 4) support covalent bond between H247 and the C5 unit moiety, the transformation from S97-bound moiety **3** to D-OH (**6**) proceeds via the following cascade: (i) The $N^{\varepsilon 2}$ atom of H247 attacks the aldehyde carbonyl of the S97-bound moiety **3** to form the hemiaminal **4**, referred to as the covalently linked intermediate molecule (CLIM). The structure of the H247- and S97-linked **4** is consistent with the electron density shape (Extended Data Fig. 2a). (ii) Lactonization of the hemiaminal **4**, via a second tetrahedron intermediate (not shown), gives free S97 and H247-bound lactone **5**. The structure of **5** is consistent with the $C_5H_5O_2$ modification of H247 detected by MS/MS (Fig. 1c, d and Extended Data Figs 3 and 4). (iii) Finally, the hydrolysis of H247-linked C5 unit will produce D-OH (**6**). A more detailed chemistry analysis is shown in the Supplementary Notes. **d**, Proposed mechanism of AtD14-mediated 4BD hydrolysis, which is similar to that of GR24.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



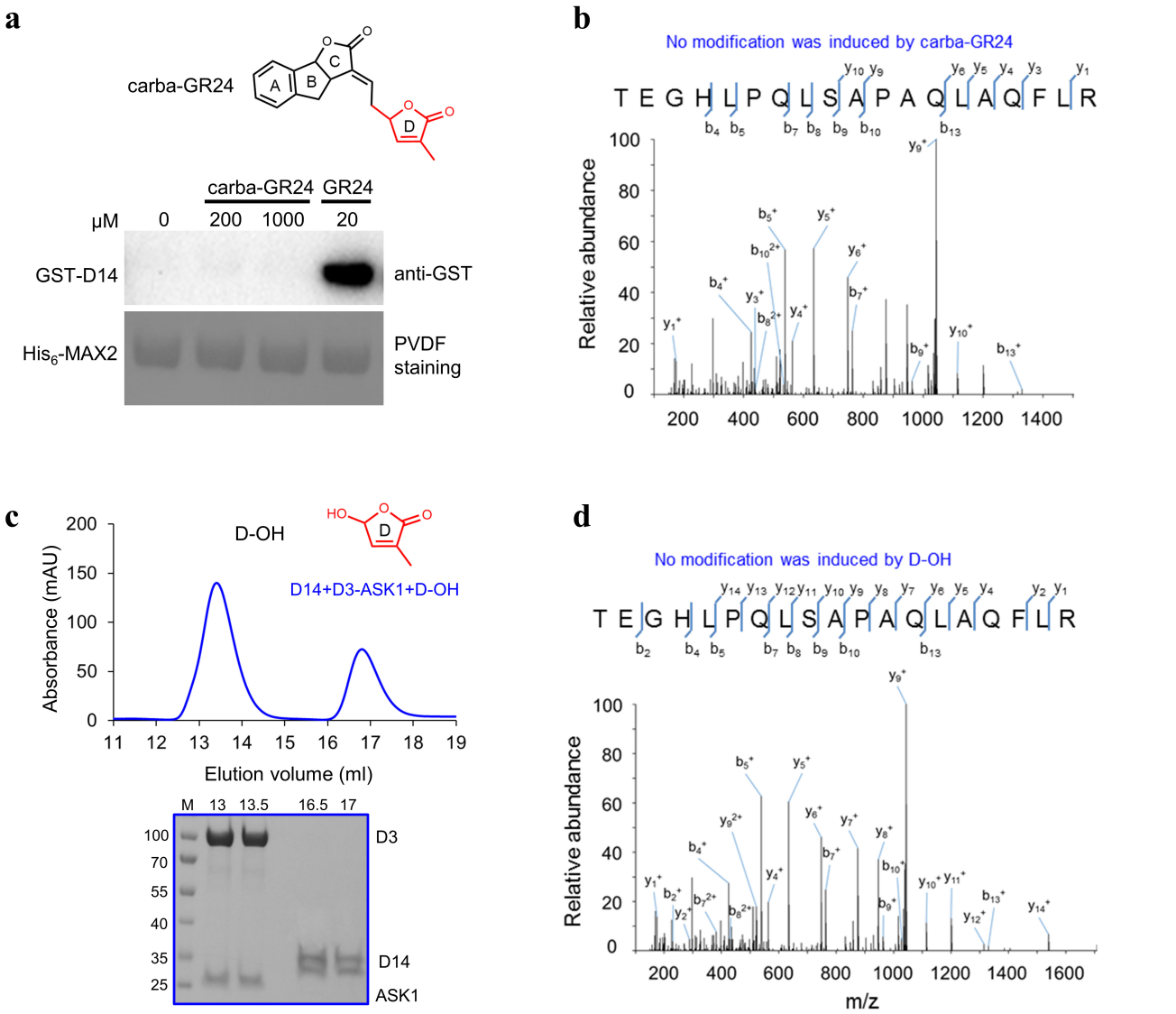
Extended Data Figure 3 | Various SL molecules generated the same $C_5H_5O_2$ modification of AtD14. a, Chemical structures of *rac*-GR24 (racemic GR24), 4BD and 5DS with the conserved D-ring. **b**, GR24-induced $C_5H_5O_2$ modification of H247 of AtD14 was identified. MS/MS spectra of a triply charged peptide (244-TEGhLPQLSAPAQLAQFLR-262) of AtD14 at mass-to-charge ratio (*m*/*z*) 725.0528 corresponding to the mass of the $C_5H_5O_2$ modification of H247. The AtD14 gel bands of the SEC-separated GR24-induced AtD14–D3 complex were excised for trypsin digestion and MS/MS analysis. Labelled peaks correspond to masses of y and b ions of the modified peptide. Lowercase 'h' indicates the modification on H247 by a GR24 hydrolysis intermediate with the molecular formula $C_5H_5O_2$. **c**, Size-exclusion chromatography analysis of the 5DS- or 4BD-induced AtD14–D3 interaction using the same method as described in Fig. 1b. **d**, **e**, 4BD (**d**) and 5DS (**e**) induced the same $C_5H_5O_2$ modification of H247, which was detected using the in-gel digestion method as described in **b**.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



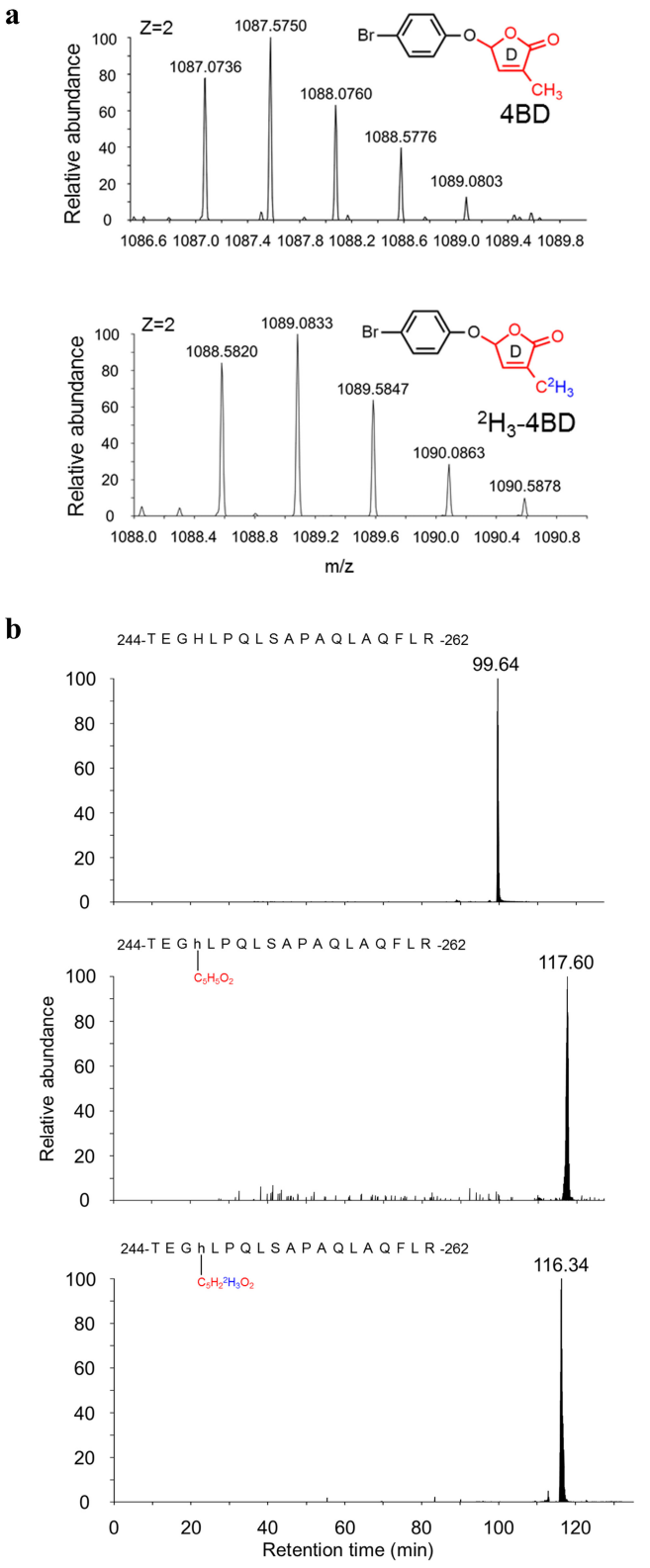
Extended Data Figure 4 | SLs generated the $C_5H_5O_2$ modification of AtD14 *in planta*. **a**, **b**, The existence of CLIM *in vivo* was suggested by successful identification of the same $C_5H_5O_2$ modification on the AtD14–Flag purified from plants pre-treated with GR24 (**a**) or 5DS (**b**). The MS/MS analysis was performed with the same method as mentioned in Extended Data Fig. 3b. **c**, No modification was observed on the 247th amino acid of AtD14(S97A). The MS/MS spectra of a triply charged peptide 244-TEGHLPQLSAPAQLAQFLR-262 of AtD14(S97A) at mass-to-charge ratio (*m*/*z*) 693.0458 corresponding to the mass of unmodified peptide, which were digested from AtD14(S97A) band separated by SDS–PAGE of the mixture containing AtD14(S97A), GR24 and D3–ASK1. **d**, No modification was observed on the 247th amino acid of AtD14(H247A). The MS/MS spectra of a triply charged peptide 244-TEGALPQLSAPAQLAQFLR-262 of AtD14(H247A) at *m*/*z* 671.0434 corresponding to the mass of unmodified peptide, which were digested from AtD14(H247A) band separated by SDS–PAGE of the mixture containing AtD14(H247A), GR24 and D3–ASK1. **e**, Size-exclusion chromatography analyses showed that the alanine substitution on S97 or H247 of AtD14 disrupted the GR24-induced interaction between the AtD14 and D3–ASK1. The size-exclusion chromatography and SDS–PAGE are performed similarly as described in Fig. 1b.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



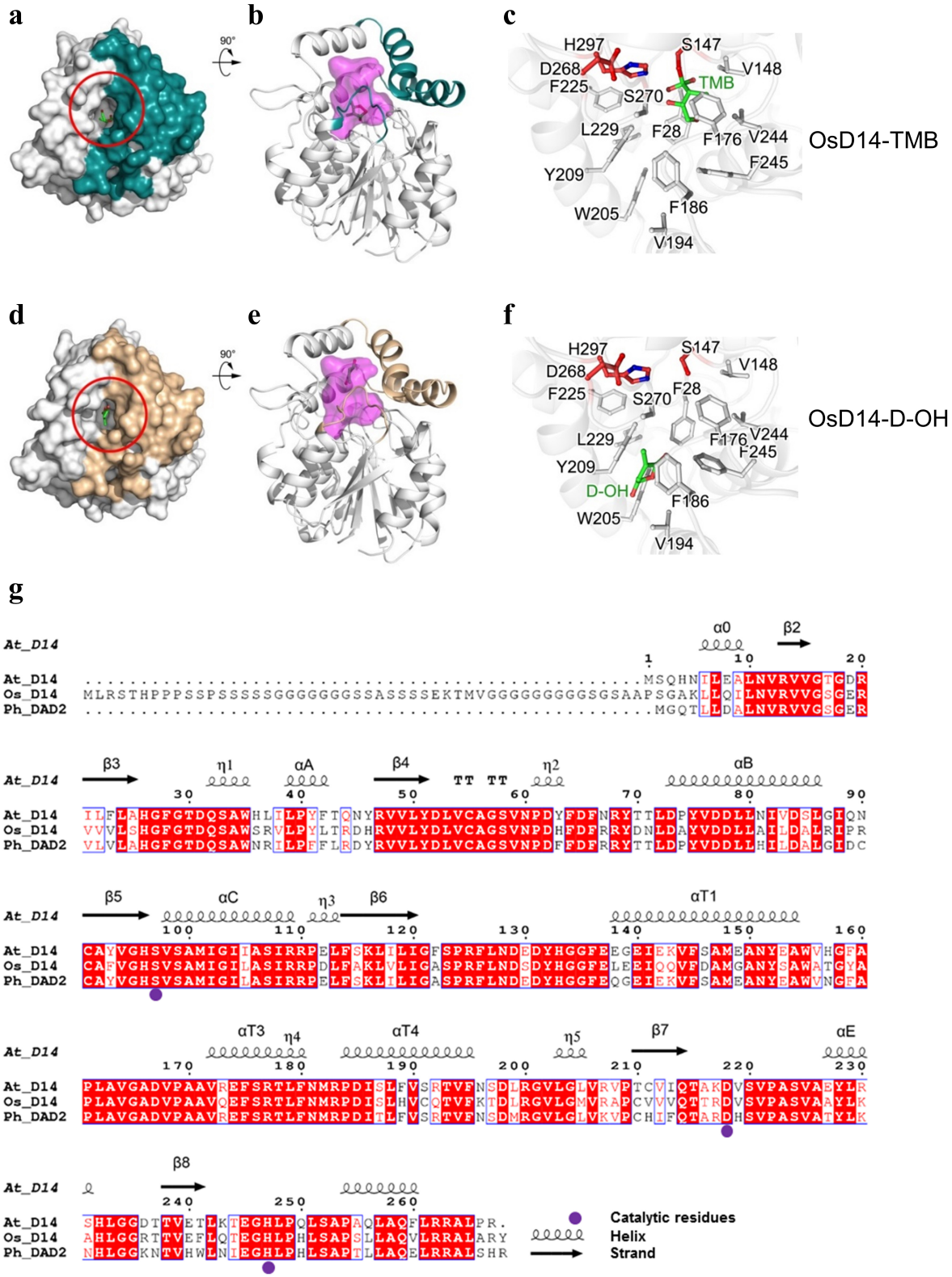
Extended Data Figure 5 | Carba-GR24 and D-OH could neither induce AtD14–D3 interaction nor generate the $C_5H_5O_2$ modification on AtD14. a, Pull-down assays using recombinant His_6–MAX2 and GST–AtD14 with indicated concentration of carba-GR24 and GR24. Full blots are shown in Supplementary Fig. 1. **b**, The MS/MS spectra of a triply charged peptide 244-TEGHLPQLSAPAQLAQFLR-262 of AtD14 at *m*/*z* of 693.0450 corresponding to the mass of an unmodified peptide, which was digested from the AtD14 bands separated by SDS–PAGE of the mixture containing AtD14, carba-GR24 and D3–ASK1. **c**, Chemically synthesized D-OH was unable to induce AtD14–D3 interaction. AtD14 was incubated with D3–ASK1 in the presence of 8 mM D-OH at 25 °C for one hour, followed by size-exclusion chromatography analysis as described in Fig. 1b. **d**, The MS/MS spectra of a triply charged peptide 244-TEGHLPQLSAPAQLAQFLR-262 of AtD14 at *m*/*z* of 693.0454 corresponding to the mass of an unmodified peptide, which was digested from the AtD14 bands separated by SDS–PAGE of the mixture containing AtD14, D-OH and D3–ASK1.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



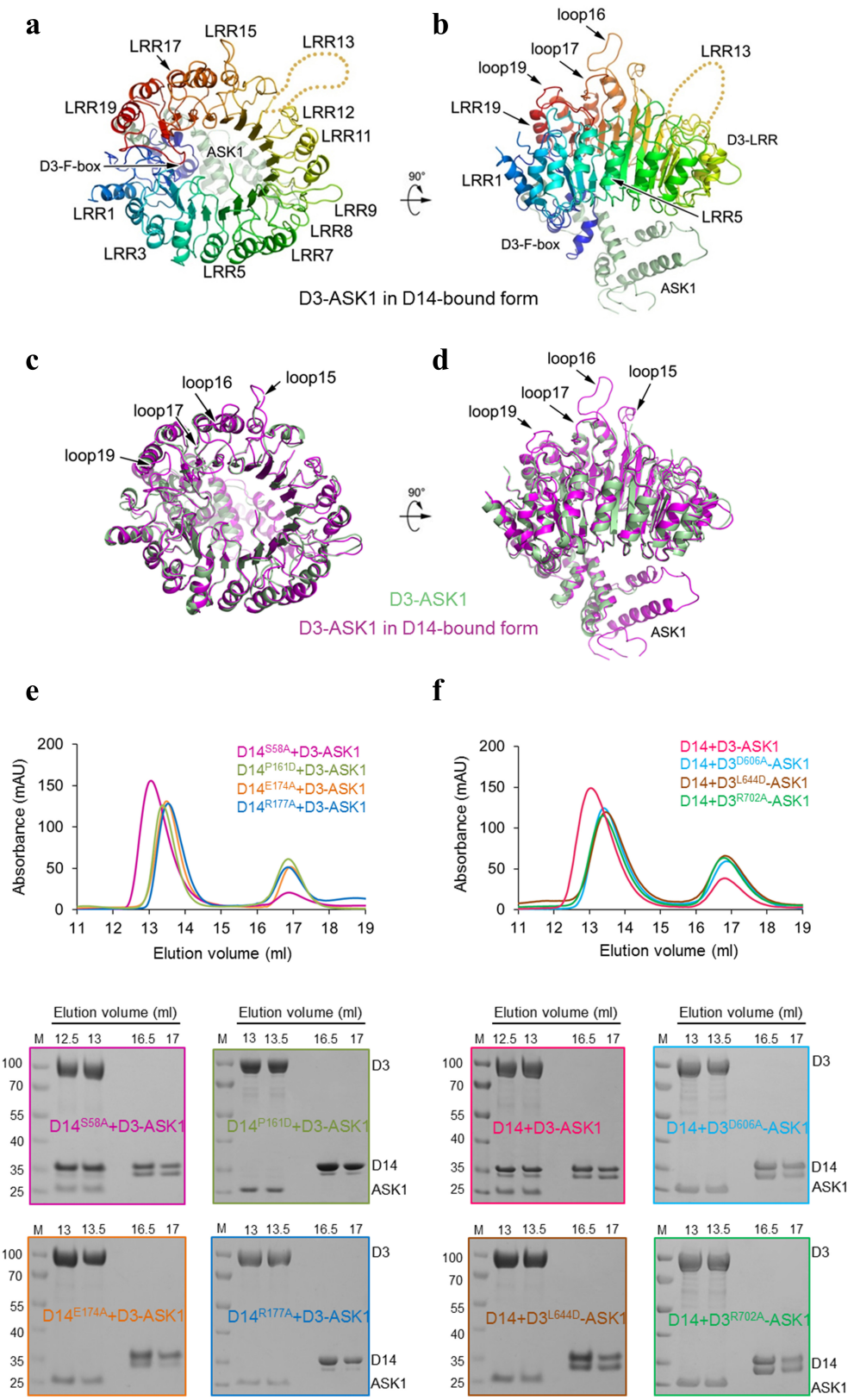
Extended Data Figure 6 | The MS spectra and chromatograms of the $C_5H_5O_2$-modified peptide of AtD14. a, The MS spectra of a bivalently charged peptide (244-TEGhLPQLSAPAQLAQFLR-262) of AtD14 at *m*/*z* of 1087.0736 or 1088.5828 corresponding to the mass of $C_5H_5O_2$ or $C_5H_2{}^2H_3O_2$ modification of the peptide, which was isolated from AtD14 in the 4BD or $^2H_3$-labelled 4BD-induced AtD14–D3 complex with in-solution digestion method, respectively. **b**, The chromatograms of the unmodified (top panel), $C_5H_5O_2$- (middle panel) or $C_5H_2{}^2H_3O_2$-modified (bottom panel) peptide of AtD14 (244-TEGHLPQLSAPAQLAQFLR-262) in the LC-MS/MS analyses. The mass spectra of the $C_5H_5O_2$- or $C_5H_2{}^2H_3O_2$-modified peptide are shown in **a**.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



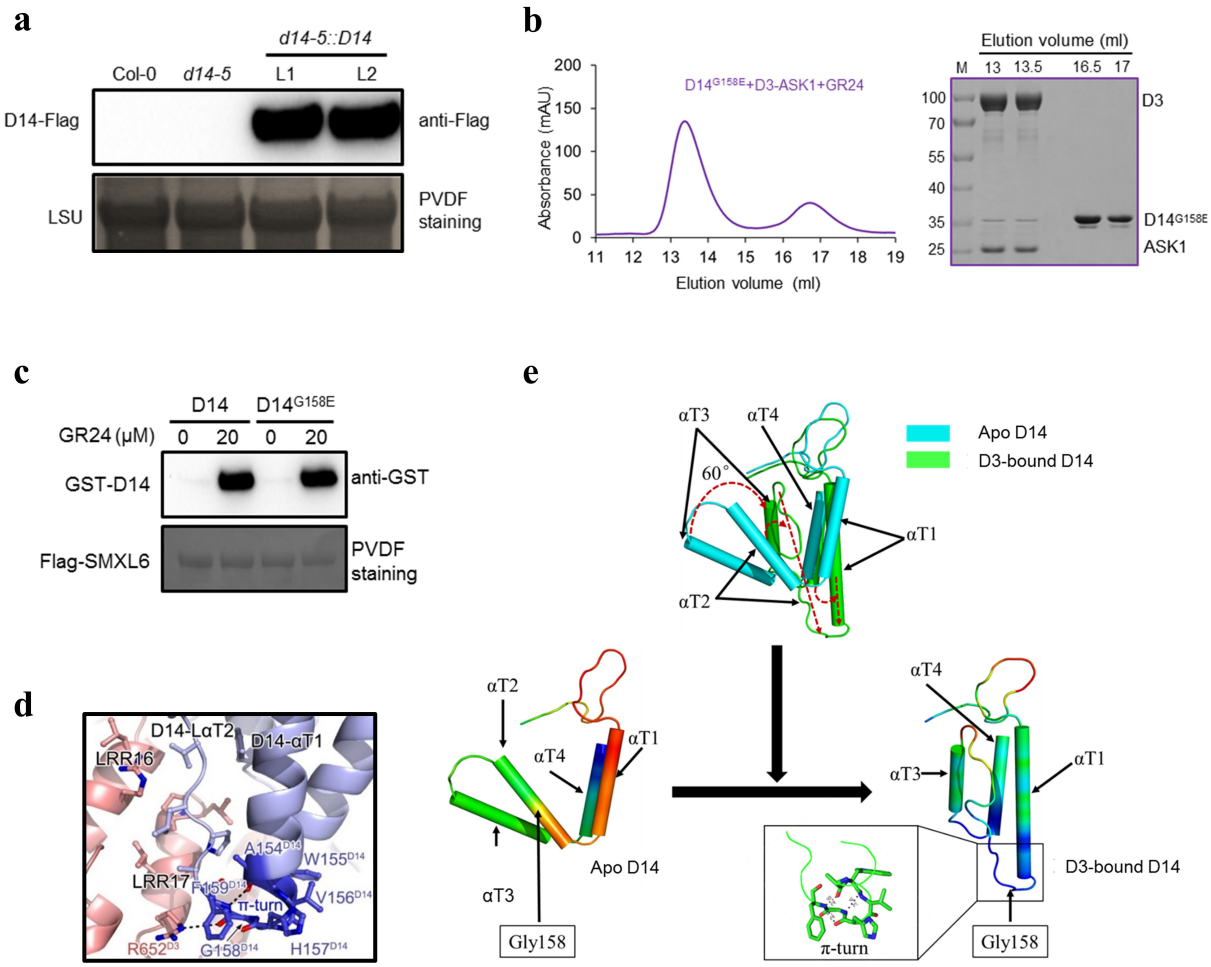
Extended Data Figure 7 | Comparison of different D14 structures. **a**–**f**, Related to Fig. 2, the characteristic features of two previously reported D3-unbound OsD14 structures, focusing on the shape (**a**, **d**), volume (**b**, **e**) and details (**c**, **f**) of the catalytic centre. (OsD14–TMB, PDB code: 4IHA; OsD14–D-OH, PDB code: 3WIO). **g**, Sequence alignment and structural annotation of D14 orthologues among different species. Secondary structure elements of AtD14 in the crystal structure of D3–AtD14 complex are displayed on top of the *Arabidopsis thaliana* D14 sequence. Identical and conserved residues are highlighted by red and white grounds, respectively. The three catalytic residues, S97, D218 and H247, are indicated by magenta dots. The GenBank accession numbers for sequences from top to bottom: *Arabidopsis thaliana* D14 (Q9SQR3), *Oryza sativa* D14 (Q10QA5) and *Petunia hybrida* DAD2 (AFR68698).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 8 | Structures of D3–ASK1 and mutagenesis analyses of the AtD14–D3 interaction. a, **b**, Overall structure of AtD14-bound D3–ASK1 in top view (**a**) and side view (**b**). ASK1 and D3 are coloured as pale green and rainbow scheme, respectively. LRRs are labelled. The unusually long loops in LRR16, LRR17 and LRR19 (denoted as loop16, loop17 and loop19) are indicated. The missing part of LRR13 is denoted as a dashed line. **c**, **d**, Structural comparison between AtD14-bound D3–ASK1 (magenta) and the apo D3–ASK1 (pale green), with their orientations related to **a** and **b**. The indicated loop15, loop16, loop17 and loop19 undergo significant conformational changes during AtD14 binding. **e**, **f**, Amino acid substitution of the central residues ($P161^{AtD14}$, $E174^{AtD14}$, $R177^{AtD14}$, $D606^{D3}$, $L644^{D3}$ and $R702^{D3}$) on the interface attenuate interaction between AtD14 and D3–ASK1. GR24 was supplemented in all the reactions. The GR24-induced interactions between D3–ASK1 and wild-type AtD14 or unrelated mutant AtD14(S58A) were used as the control. The size-exclusion chromatography and SDS–PAGE were performed as described in Fig. 1b.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 9 | Gly158 plays an important role in the SL-induced open-to-closed transition of AtD14 lid. a, AtD14–Flag protein levels in two independent lines (L1 and L2) of the *d14-5* plants with transgenic expression of Flag-tagged AtD14 (*d14-5::AtD14*, mentioned in Fig. 4b) (upper panel). Staining of Rubisco large subunit (LSU) served as a loading control (lower panel). Full blots are shown in Supplementary Fig. 1. **b**, AtD14(G158E) was incubated with D3–ASK1 in the presence of GR24, followed by size-exclusion chromatography analysis as described in Fig. 1b. **c**, AtD14(G158E) maintains the SL-induced interaction with SMXL6. Pull-down assays using recombinant SMXL6–Flag and GST–AtD14 or GST–AtD14(G158E) in the absence or presence of 20 μM GR24. Full blots are shown in Supplementary Fig. 1. **d**, $G158^{AtD14}$ participates in the construction of a π-turn structure and the formation of a hydrogen bond with $R652^{D3}$. The π-turn structure, and structural elements of AtD14 and D3 are coloured as blue, light blue and salmon, respectively. The π-turn residues include A154, W155, V156, H157, G158 and F159. **e**, During the open (apo AtD14, cyan) to closed (D3-bound AtD14, green) state transition of AtD14, the αT2 and αT3 helices of the lid domain experience pronounced structural and positional alterations (upper panel). In the open state (lower-left panel), the N terminus of αT2 and portions from αT1 and its immediate preceding loop are more flexible than other parts of the AtD14 lid domain, as judged by the higher temperature factors (B-factors) of these regions. In the closed state (lower-right panel), the π-turn structure, with an intrinsic capacity to stop further extension of the αT1 C terminus during the open-to-closed transition of AtD14, contributes to stabilizing proper conformation of the AtD14 lid in the AtD14–D3 complex. Therefore, substitution of Gly158 with a good helix-forming residue Glu would expectedly impair normal conformational changes of the lid which was observed in the AtD14–D3 complex.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 1 | Data collection and refinement statistics

	D14-D3-ASK1 complex		D3-ASK1
	Se-SAD peak	Native	
PDB codes (*)		**5HZG**	**5HYW**
Data collection statistics			
Cell parameters			
a (Å)	107.1	106.4	79.7
b (Å)	173.2	173.0	79.7
c (Å)	187.4	186.8	327.3
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Space group	$P2_12_12_1$	$P2_12_12_1$	$P4_3$
Wavelength used (Å)	0.9792	0.9792	0.9792
Resolution (Å)	50.0–3.6 (3.7-3.6)[c]	50.0–3.3 (3.4-3.3)	50.0-3.0 (3.05-3.0)
No. of all reflections	594,727 (41,118)	231,240 (50,269)	141,268 (7,085)
No. of unique reflections	58,980 (4,041)	11,666 (3,153)	38,298 (1,969)
Completeness (%)	100.0 (100.0)	99.9 (99.9)	94.1 (95.5)
Average $I/\sigma(I)$	31.6 (6.5)	12.6 (3.1)	12.0 (2.8)
R_{merge}[a] (%)	15.2 (64.6)	11.3 (50.3)	14.0 (54.6)
Refinement statistics			
No. of reflections used ($\sigma(F) > 0$)		49,674	38,237
R_{work}[b] (%)		24.6	23.5
R_{free}[b] (%)		31.6	31.2
r.m.s.d. bond distance (Å)		0.011	0.011
r.m.s.d. bond angle (°)		1.781	1.671
Average B-value (Å^2)		89.0	62.0
No. of atoms		15,676	10,174
Ramachandran plot			
Res. in favored regions (%)		100.0	100.0
Res. in outlier regions (%)		0.0	0.0

[a] $R_{merge} = \Sigma_h\Sigma_i|I_{h,i} - I_h|/\Sigma_h\Sigma_i I_{h,i}$, where I_h is the mean intensity of the i observations of symmetry related reflections of h.
[b] $R_{work} = \sum(||F_p(obs)| - |F_p(calc)||)/\sum|F_p(obs)|$; R_{free} is an R factor for a pre-selected subset (5%) of reflections that was not included in refinement. F_p, structure factor of protein.
[c] Numbers in parentheses are corresponding values for the highest resolution shell.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

LETTER

doi:10.1038/nature18952

Vaccine protection against Zika virus from Brazil

Rafael A. Larocca[1]*, Peter Abbink[1]*, Jean Pierre S. Peron[2], Paolo M. de A. Zanotto[2], M. Justin Iampietro[1], Alexander Badamchi-Zadeh[1], Michael Boyd[1], David Ng'ang'a[1], Marinela Kirilova[1], Ramya Nityanandam[1], Noe B. Mercado[1], Zhenfeng Li[1], Edward T. Moseley[1], Christine A. Bricault[1], Erica N. Borducchi[1], Patricia B. Giglio[1], David Jetton[1], George Neubauer[1], Joseph P. Nkolola[1], Lori F. Maxfield[1], Rafael A. De La Barrera[3], Richard G. Jarman[3], Kenneth H. Eckels[3], Nelson L. Michael[3], Stephen J. Thomas[3] & Dan H. Barouch[1,4]

Zika virus (ZIKV) is a flavivirus that is responsible for the current epidemic in Brazil and the Americas[1,2]. ZIKV has been causally associated with fetal microcephaly, intrauterine growth restriction, and other birth defects in both humans[3–8] and mice[9–11]. The rapid development of a safe and effective ZIKV vaccine is a global health priority[1,2], but very little is currently known about ZIKV immunology and mechanisms of immune protection. Here we show that a single immunization with a plasmid DNA vaccine or a purified inactivated virus vaccine provides complete protection in susceptible mice against challenge with a strain of ZIKV involved in the outbreak in northeast Brazil. This ZIKV strain has recently been shown to cross the placenta and to induce fetal microcephaly and other congenital malformations in mice[11]. We produced DNA vaccines expressing ZIKV pre-membrane and envelope (prM-Env), as well as a series of deletion mutants. The prM-Env DNA vaccine, but not the deletion mutants, afforded complete protection against ZIKV, as measured by absence of detectable viraemia following challenge, and protective efficacy correlated with Env-specific antibody titers. Adoptive transfer of purified IgG from vaccinated mice conferred passive protection, and depletion of CD4 and CD8 T lymphocytes in vaccinated mice did not abrogate this protection. These data demonstrate that protection against ZIKV challenge can be achieved by single-shot subunit and inactivated virus vaccines in mice and that Env-specific antibody titers represent key immunologic correlates of protection. Our findings suggest that the development of a ZIKV vaccine for humans is likely to be achievable.

The World Health Organization declared the clusters of microcephaly and neurological disorders and their association with ZIKV infection to be a global public health emergency on February 1, 2016. ZIKV is believed to cause neuropathology in developing fetuses by crossing the placenta and targeting cortical neural progenitor cells[9–14], leading to impaired neurogenesis and resulting in microcephaly and other congenital malformations. ZIKV has also been associated with neurologic conditions in adults, such as Guillain-Barré syndrome[15].

Vaccines have been developed for other flaviviruses, including yellow fever virus, Japanese encephalitis virus, tick-borne encephalitis virus, and dengue viruses, but no vaccine currently exists for ZIKV. To develop preclinical challenge models for candidate ZIKV vaccines, we obtained low-passage ZIKV isolates from northeast Brazil (Brazil/ZKV2015; University of São Paulo)[11] and Puerto Rico (PRVABC59; US Centers for Disease Control and Prevention) (Extended Data Fig. 1). We expanded these viruses in Vero cells to generate preclinical challenge stocks, which we termed ZIKV-BR and ZIKV-PR, respectively. These ZIKV strains are part of the Asian ZIKV lineage[16] and differ from each other by five amino acids in the polyprotein (Extended Data Fig. 2). The Brazil/ZKV2015 strain has also recently been reported to recapitulate key clinical manifestations, including fetal microcephaly and intrauterine growth restriction, in wild-type SJL mice[11]. Similarly, the related French Polynesian H/PF/2013 strain has been shown to induce placental damage and fetal demise in *Ifnar*$^{-/-}$ C57BL/6 mice as well as in wild-type C57BL/6 mice following IFN-α receptor blockade[10].

We designed ZIKV prM-Env immunogens based on the Brazil BeH815744 strain (Extended Data Fig. 2) and optimized them for increased antigen expression. We also designed deletion mutants lacking prM and/or lacking the transmembrane region (ΔTM) or the full stem (Δstem) of Env (Fig. 1a). Plasmid DNA vaccines encoding these antigens were produced, and transgene expression was verified by western blot (Fig. 1b). To assess the immunogenicity of these vaccines, groups of Balb/c mice ($n = 5$–10 per group) received a single immunization of 50 μg of each DNA vaccine by the intramuscular (i.m.) route at week 0. Env-specific antibody responses were evaluated at week 3 by ELISA. The prM-Env DNA vaccine elicited higher Env-specific antibody titers than did the Env DNA vaccine and all of the ΔTM and Δstem deletion mutants (Fig. 1c), indicating the importance of including prM as well as the full-length Env sequence. No prM-specific antibody responses were detected (Extended Data Fig. 3). The prM-Env DNA vaccine also induced ZIKV-specific neutralizing antibodies after a single immunization (Table 1), as measured by a virus-specific microneutralization assay[17]. In addition, the prM-Env DNA vaccine induced Env-specific $CD8^+$ and $CD4^+$ T-lymphocyte responses, as assessed by IFNγ ELISPOT and multiparameter intracellular cytokine staining assays (Fig. 1d, e).

To assess the protective efficacy of these DNA vaccines against ZIKV challenge, we infected vaccinated or sham control Balb/c mice at week 4 by the intravenous (i.v.) route with 10^5 viral particles (10^2 plaque-forming units (PFU)) of ZIKV-BR or ZIKV-PR. Viral loads following ZIKV challenge were quantitated by RT–PCR[18]. Sham-vaccinated mice inoculated with ZIKV-BR developed approximately 6 days of detectable viraemia with a mean peak viral load of 5.42 log copies per ml (range 4.55–6.57 log copies per ml; $n = 10$) on day 3 after challenge (Fig. 2a). In contrast, a single immunization with the prM-Env DNA vaccine provided complete protection against ZIKV-BR challenge with no detectable viraemia (<100 copies per ml) at any time point ($n = 10$). Complete protection was also observed when vaccinated mice were challenged at week 8 (data not shown). The prM-Env DNA vaccine also afforded complete protection against ZIKV-PR challenge ($n = 5$). ZIKV-PR replicated to slightly lower levels (mean peak viral load 4.96 log copies per ml; range 4.80–5.33 log copies per ml; $n = 5$) than did ZIKV-BR in sham controls. In contrast with the prM-Env DNA vaccine, the DNA vaccines lacking prM as well as the ΔTM and Δstem deletion mutants did not provide complete protection against ZIKV-BR challenge, although viral loads were still reduced in these animals as compared with sham controls (Fig. 2b).

The varying degrees of protection obtained with this set of DNA vaccines allowed for an analysis of immune correlates of protection. Protective efficacy correlated with Env-specific binding antibody titers

[1]Center for Virology and Vaccine Research, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts 02215, USA. [2]University of São Paulo, São Paulo 05508-000, Brazil. [3]Walter Reed Army Institute of Research, Silver Spring, Maryland 20910, USA. [4]Ragon Institute of MGH, MIT, and Harvard, Cambridge, Massachusetts 02139, USA.
*These authors contributed equally to this work.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

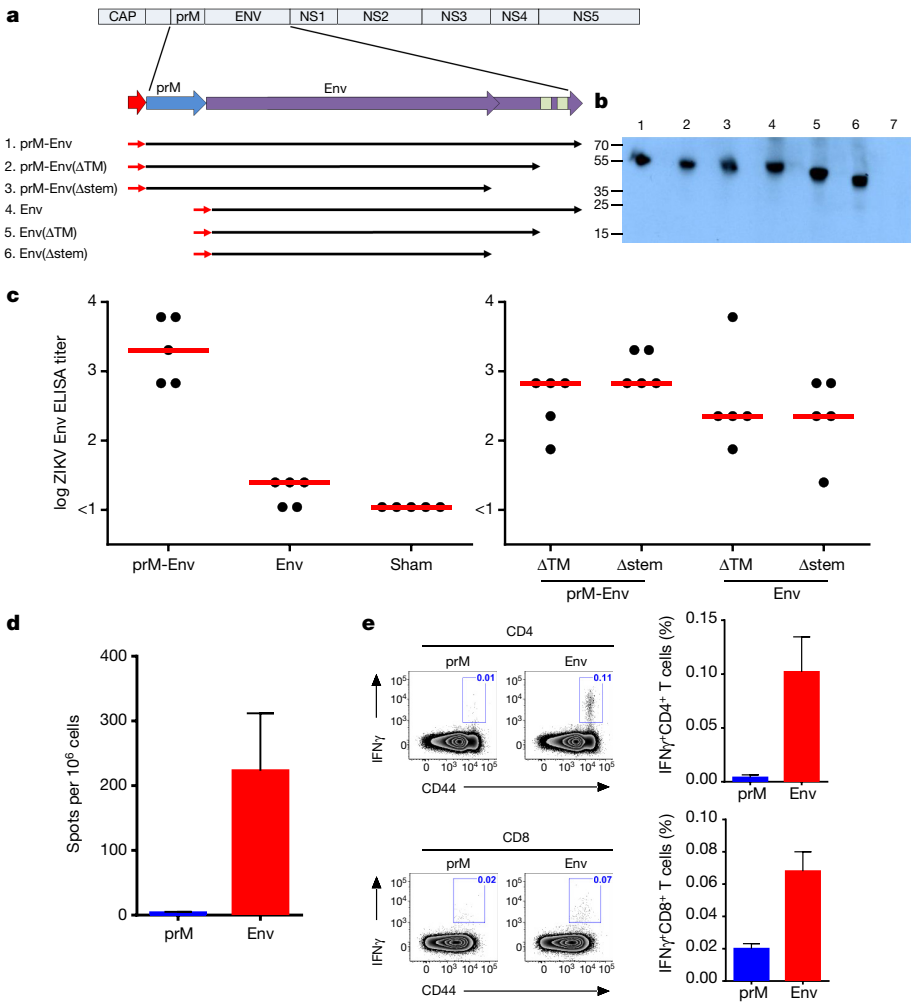


Figure 1 | Construction and immunogenicity of DNA vaccines. a, Schema of ZIKV prM-Env immunogens and deletion mutants. **b**, Western blot of transgene expression from (1) prM-Env, (2) prM-Env(ΔTM), (3) prM-Env(Δstem), (4) Env, (5) Env(ΔTM), (6) Env(Δstem), and (7) sham DNA vaccines transfected in 293T cells. Balb/c mice ($n = 5$ per group) received a single immunization with 50 μg of these DNA vaccines by the i.m. route. **c**, Humoral immune responses were assessed at week 3 following vaccination by Env-specific ELISA. Red bars reflect medians. **d**, **e**, Cellular immune responses were assessed by IFNγ ELISPOT assays (**d**) and multi-parameter intracellular cytokine staining assays (**e**). Error bars reflect s.e.m.

($P = 0.0005$ comparing protected versus infected animals; Fig. 2c) as well as ZIKV-specific neutralizing antibody titers >10 (Table 1). In addition, peak viral loads on day 3 were inversely correlated with antibody titers ($P < 0.0001$, $R^2 = 0.72$; Fig. 2d). These data suggest that Env-specific antibodies were critical for the protective efficacy of DNA vaccines against ZIKV-BR challenge. Mice that received two immunizations with the prM-Env DNA vaccine at week 0 and week 4 developed high neutralizing antibody titers of 1,022 at week 8 (Table 1) and were also protected against ZIKV-BR challenge (data not shown).

The prM-Env DNA vaccine also provided complete protection against ZIKV-BR challenge in SJL mice (Extended Data Fig. 4) and against both ZIKV-BR and ZIKV-PR challenge in C57BL/6 mice (Extended Data Figs 5 and 6). ZIKV-BR replicated efficiently in SJL mice, consistent with a previous study[11], although at slightly lower levels (mean peak viral load 4.70 log copies per ml; range 3.50–5.92 log copies per ml; $n = 5$) than in Balb/c mice (Fig. 2a). In contrast, both ZIKV-BR and ZIKV-PR replicated poorly in C57BL/6 mice (Extended Data Fig. 5), also consistent with previous reports, potentially as a result of robust IFN-α-mediated innate immune restriction in this strain of mice[10,11,19,20].

To investigate the immunological mechanism of protection against ZIKV-BR challenge, we purified IgG from serum from Balb/c mice vaccinated with prM-Env DNA. Passive infusion of varying quantities of purified IgG by the i.v. route resulted in median Env-specific log serum antibody titers of 2.82 (high), 2.35 (mid) and 1.87 (low) in recipient mice following adoptive transfer (Fig. 3a). All recipient mice with log serum antibody titers of 2.35 or higher were protected against ZIKV-BR challenge (Fig. 3b, c), demonstrating that protection can be mediated by vaccine-elicited IgG alone and confirming that the magnitude of Env-specific antibody titers correlates with protective efficacy ($P < 0.0001$, Fig. 3b). In contrast, only 1 out of 5 recipient mice that received low levels of Env-specific IgG were protected, although they still exhibited reduced viral loads compared with sham controls (Extended Data Fig. 7). These data define the minimum threshold of Env-specific antibody titers required for protection in this model.

We next depleted $CD4^+$ and/or $CD8^+$ T lymphocytes in prM-Env-vaccinated mice on day −2 and day −1 before challenge (>99.9% efficiency; Extended Data Fig. 8). Depletion of these T-lymphocyte subsets did not detectably abrogate the protective efficacy of the prM-Env DNA vaccine against ZIKV-BR challenge (Fig. 3d). These data indicate that Env-specific T-lymphocyte responses were not required for protection in this model, although these findings do not exclude the possibility that ZIKV-specific cellular immune responses may be beneficial in other settings.

To extend these observations to a vaccine platform that has historically provided clinical efficacy against other flaviviruses, we explored the immunogenicity and protective efficacy of a ZIKV purified inactivated virus (PIV) vaccine derived from the Puerto Rico PRVABC59

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

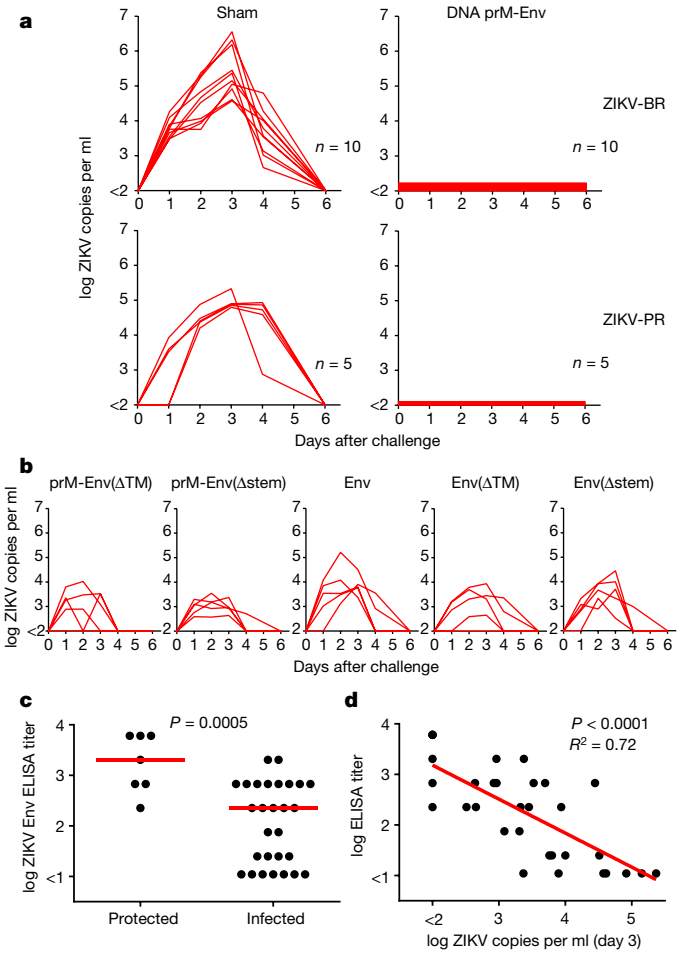


Figure 2 | Protective efficacy of DNA vaccines. a, Balb/c mice ($n = 5$ or 10 per group) received a single immunization by the i.m. route with 50 μg prM-Env DNA vaccine or a sham vaccine and were challenged at week 4 by the i.v. route with 10^5 viral particles (10^2 PFU) ZIKV-BR or ZIKV-PR. Serum viral loads are shown. **b**, Mice ($n = 5$ per group) received a single immunization with 50 μg of various DNA vaccines and were challenged with ZIKV-BR. **c**, **d**, Correlates of protective efficacy (**c**) and day 3 viral loads (**d**) are shown. Red bars reflect medians. *P* values and R^2 values reflect *t*-tests and Spearman rank-correlation tests.

strain. Groups of Balb/c mice ($n = 5$ per group) received a single immunization of 1 μg of the PIV vaccine with alum or alum alone by the i.m. or subcutaneous (s.c.) routes. Antibody titers were higher in the group that received the PIV vaccine by the i.m. route rather than

Table 1 | ZIKV-specific neutralizing antibody titers

Vaccine	ZIKV MN50 titer
DNA prM-Env	22
DNA prM-Env(ΔTM)	<10
DNA prM-Env(Δstem)	<10
DNA Env	<10
DNA Env(ΔTM)	<10
DNA Env(Δstem)	<10
DNA prM-Env + boost	1,022
PIV + alum i.m.	15
PIV + alum s.c.	15
Sham + alum i.m.	<10
Sham + alum s.c.	<10
Anti-flavivirus antibody	232

Balb/c mice received a single immunization with 50 μg of various DNA vaccines (Figs 1, 2) or 1 μg purified inactivated virus (PIV) vaccines with alum (Fig. 4), and pooled serum was assessed for ZIKV-specific neutralizing antibodies at week 4. 50% microneutralization (MN50) titers are shown. Also shown are MN50 titers in serum from mice following two immunizations with DNA-prM-Env (boost) and an anti-flavivirus human polyclonal antibody.

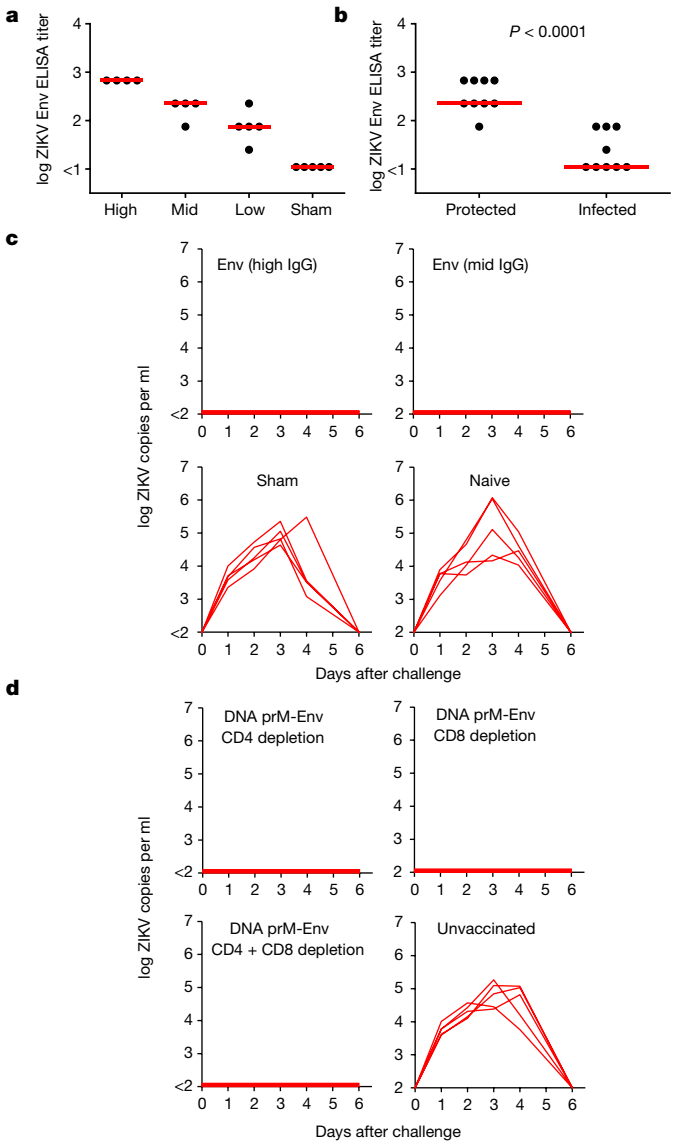


Figure 3 | Mechanistic studies. a, Env-specific serum antibody titers in recipient Balb/c mice ($n = 5$ per group) following adoptive transfer of varying amounts (high, mid, low) of IgG purified from serum from mice vaccinated with prM-Env DNA or naive mice (sham). **b**, Correlates of protective efficacy. **c**, Serum viral loads in mice that received adoptive transfer of purified IgG from vaccinated mice and were challenged with ZIKV-BR. **d**, Serum viral loads in prM-Env-DNA-vaccinated mice that were depleted of $CD4^+$ and/or $CD8^+$ T lymphocytes before challenge with ZIKV-BR. Red bars reflect medians. *P* values reflect *t*-tests.

by the s.c. route, as compared by ELISA (Fig. 4a). The PIV vaccine by both routes also induced ZIKV-specific neutralizing antibodies after a single immunization (Table 1). At week 4, all mice were i.v. challenged with ZIKV-BR as described above. Complete protection was observed in the group that received the PIV vaccine by the i.m. route (Fig. 4b, c). Two mice that received the PIV vaccine by the s.c. route showed brief low levels of viraemia (Fig. 4c), potentially consistent with the lower Env-specific binding antibody titers in this group (Fig. 4b).

Our data demonstrate that a single immunization with a DNA vaccine or a PIV vaccine provided complete protection against parenteral ZIKV challenge in mice. The prM-Env DNA vaccine afforded protection in three strains of mice and against both ZIKV-BR and ZIKV-PR challenges, suggesting the generalizability of these observations. Protective efficacy was mediated by vaccine-elicited Env-specific antibodies, as evidenced by (1) statistical analyses of immune correlates

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

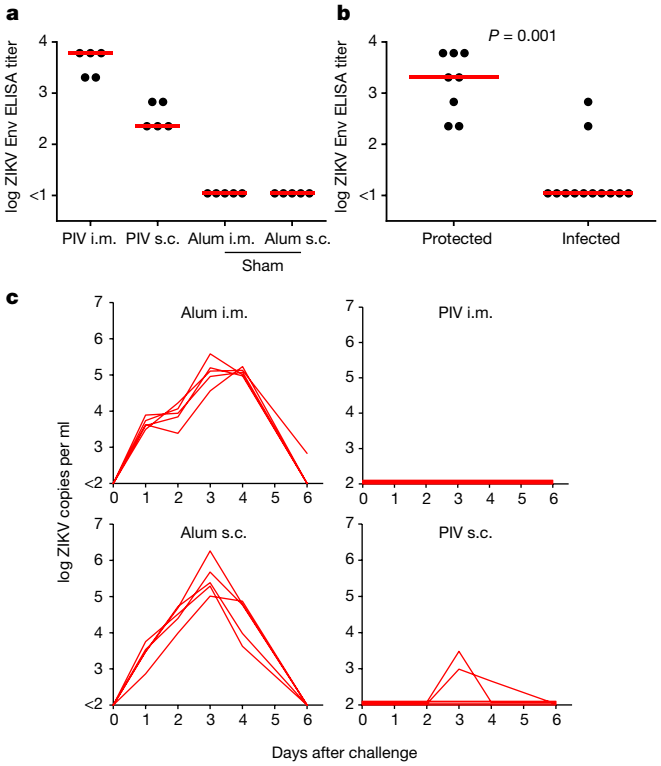


Figure 4 | Immunogenicity and protective efficacy of the PIV vaccine. Balb/c mice ($n = 5$ per group) received a single immunization by the i.m. or s.c. route with 1 μg PIV vaccine with alum, or alum alone, and were challenged at week 4 by the i.v. route with 10^5 viral particles (10^2 PFU) ZIKV-BR. **a**, Humoral immune responses were assessed at week 3 following vaccination by Env-specific ELISA. **b**, Correlates of protective efficacy. **c**, Serum viral loads are shown following ZIKV-BR challenge. Red bars reflect medians. *P* values reflect *t*-tests.

of protection (Figs 2c, d, 4b), (2) adoptive transfer studies with purified IgG from vaccinated mice (Fig. 3a–c), and (3) T-lymphocyte depletion studies in vaccinated mice (Fig. 3d). The adoptive transfer studies also defined the threshold of Env-specific antibody titers required for protection in this model.

It is difficult to extrapolate directly the results from these vaccine studies in mice to potential clinical efficacy in humans. Nevertheless, the robust protection observed in the present studies and the clear immune correlates of protection suggest a path forward for ZIKV vaccine development in humans. Of note, similar antibody-based correlates of protection, including neutralizing antibody titers >10, have been reported for other flavivirus vaccines, including yellow fever virus, tick-borne encephalitis virus, and Japanese encephalitis virus[21–23]. Moreover, the ZIKV-BR challenge isolate used in the present study has been shown in wild-type SJL mice to recapitulate certain key clinical findings of ZIKV infection in humans, including fetal microcephaly and intrauterine growth retardation[11]. ZIKV-BR did not lead to a fatal outcome in wild-type Balb/c and SJL mice, as has been observed in *Ifnar*$^{-/-}$ C57BL/6 mice[10,19,20], but the magnitude and duration of viraemia in Balb/c and SJL mice appear comparable with that in humans[2], suggesting the potential relevance of this model. It is notable that ZIKV-BR replicated efficiently in wild-type Balb/c and SJL mice (Fig. 2a, Extended Data Fig. 4), but replicated poorly in wild-type C57BL/6 mice (Extended Data Fig. 5), which is consistent with previous observations[10,11] and indicates important strain-specific differences for ZIKV infectivity. Further investigation into the immunologic mechanisms underlying these differences may lead to insights into innate immune control of ZIKV. Moreover, further characterization of the susceptible Balb/c and SJL murine models may facilitate future studies of ZIKV pathogenesis and the development of antiviral interventions. Future studies will also need to address the potential relevance of cross-reactive antibodies against dengue virus and other flaviviruses on ZIKV vaccine immunogenicity and protective efficacy.

The epidemiology of the current ZIKV outbreak[1,2] and the clinical consequences for fetuses in pregnant women who become infected[3–8] necessitate the urgent development of a ZIKV vaccine. Our data demonstrate that complete protection against ZIKV challenge was reliably and robustly achieved with both DNA vaccines and purified inactivated virus vaccines in susceptible mice. These vaccine platforms have previously been used at comparable doses to develop vaccines for other flaviviruses, including West Nile virus[24,25], dengue viruses[26,27], tick-borne encephalitis virus[28,29], and Japanese encephalitis virus[30], and may offer safety advantages over live attenuated and replicating flavivirus vaccines, particularly for pregnant women. Moreover, the magnitude of Env-specific antibody titers that provide complete protection against ZIKV challenge in mice should be readily achievable by DNA vaccines and purified inactivated virus vaccines in humans. Taken together, our findings provide substantial optimism that the development of a safe and effective ZIKV vaccine for humans will probably be feasible.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 30 May; accepted 22 June 2016.
Published online 28 June 2016.

1. Fauci, A. S. & Morens, D. M. Zika virus in the Americas—yet another arbovirus threat. *N. Engl. J. Med.* **374,** 601–604 (2016).
2. Petersen, L. R., Jamieson, D. J., Powers, A. M. & Honein, M. A. Zika virus. *N. Engl. J. Med.* **374,** 1552–1563 (2016).
3. Mlakar, J. *et al.* Zika virus associated with microcephaly. *N. Engl. J. Med.* **374,** 951–958 (2016).
4. Calvet, G. *et al.* Detection and sequencing of Zika virus from amniotic fluid of fetuses with microcephaly in Brazil: a case study. *Lancet Infect. Dis.* **16,** 653–660 (2016).
5. Brasil, P. *et al.* Zika virus infection in pregnant women in Rio de Janeiro — preliminary report. *N. Engl. J. Med.* http://dx.doi.org/10.1056/NEJMoa1602412 (2016).
6. Driggers, R. W. *et al.* Zika virus infection with prolonged maternal viremia and fetal brain abnormalities. *N. Engl. J. Med.* **374,** 2142–2151 (2016).
7. Rasmussen, S. A., Jamieson, D. J., Honein, M. A. & Petersen, L. R. Zika virus and birth defects—reviewing the evidence for causality. *N. Engl. J. Med.* **374,** 1981–1987 (2016).
8. Johansson, M. A., Mier-Y-Teran-Romero, L., Reefhuis, J., Gilboa, S. M. & Hills, S. L. Zika and the risk of microcephaly. *N. Engl. J. Med.* **375,** 1–4 (2016).
9. Li, C. *et al.* Zika virus disrupts neural progenitor development and leads to microcephaly in mice. *Cell Stem Cell* **19,** 120–126 (2016).
10. Miner, J. J. *et al.* Zika virus infection during pregnancy in mice causes placental damage and fetal demise. *Cell* **165,** 1081–1091 (2016).
11. Cugola, F. R. *et al.* The Brazilian Zika virus strain causes birth defects in experimental models. *Nature* **534,** 267–271 (2016).
12. Garcez, P. P. *et al.* Zika virus impairs growth in human neurospheres and brain organoids. *Science* **352,** 816–818 (2016).
13. Qian, X. *et al.* Brain-region-specific organoids using mini-bioreactors for modeling ZIKV exposure. *Cell* **165,** 1238–1254 (2016).
14. Tang, H. *et al.* Zika virus infects human cortical neural progenitors and attenuates their growth. *Cell Stem Cell* **18,** 587–590 (2016).
15. Brasil, P. *et al.* Guillain-Barré syndrome associated with Zika virus infection. *Lancet* **387,** 1482 (2016).
16. Faria, N. R. *et al.* Zika virus in the Americas: Early epidemiological and genetic findings. *Science* **352,** 345–349 (2016).
17. Thomas, S. J. *et al.* A phase II, randomized, safety and immunogenicity study of a re-derived, live-attenuated dengue virus vaccine in healthy adults. *Am. J. Trop. Med. Hyg.* **88,** 73–88 (2013).
18. Tartaglia, L. J. *et al.* Production of mucosally transmissible SHIV challenge stocks from HIV-1 circulating recombinant form 01_AE *env* sequences. *PLoS Pathog.* **12,** e1005431 (2016).
19. Lazear, H. M. *et al.* A mouse model of Zika virus pathogenesis. *Cell Host Microbe* **19,** 720–730 (2016).
20. Rossi, S. L. *et al.* Characterization of a novel murine model to study Zika virus. *Am. J. Trop. Med. Hyg.* **94,** 1362–1369 (2016).
21. Hombach, J., Solomon, T., Kurane, I., Jacobson, J. & Wood, D. Report on a WHO consultation on immunological endpoints for evaluation of new Japanese encephalitis vaccines, WHO, Geneva, 2–3 September, 2004. *Vaccine* **23,** 5205–5211 (2005).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

22. Kreil, T. R., Burger, I., Bachmann, M., Fraiss, S. & Eibl, M. M. Antibodies protect mice against challenge with tick-borne encephalitis virus (TBEV)-infected macrophages. *Clin. Exp. Immunol.* **110,** 358–361 (1997).
23. Mason, R. A., Tauraso, N. M., Spertzel, R. O. & Ginn, R. K. Yellow fever vaccine: direct challenge of monkeys given graded doses of 17D vaccine. *Appl. Microbiol.* **25,** 539–544 (1973).
24. Martin, J. E. *et al.* A West Nile virus DNA vaccine induces neutralizing antibody in healthy adults during a phase 1 clinical trial. *J. Infect. Dis.* **196,** 1732–1740 (2007).
25. Ledgerwood, J. E. *et al.* A West Nile virus DNA vaccine utilizing a modified promoter induces neutralizing antibody in younger and older healthy adults in a phase I clinical trial. *J. Infect. Dis.* **203,** 1396–1404 (2011).
26. Martinez, L. J. *et al.* Safety and immunogenicity of a dengue virus serotype-1 purified-inactivated vaccine: results of a phase 1 clinical trial. *Am. J. Trop. Med. Hyg.* **93,** 454–460 (2015).
27. Fernandez, S. *et al.* An adjuvanted, tetravalent dengue virus purified inactivated vaccine candidate induces long-lasting and protective antibody responses against dengue challenge in rhesus macaques. *Am. J. Trop. Med. Hyg.* **92,** 698–708 (2015).
28. Demicheli, V., Debalini, M. G. & Rivetti, A. Vaccines for preventing tick-borne encephalitis. *Cochrane Database Syst. Rev.* CD000977 (2009).
29. Demicheli, V., Graves, P., Pratt, M. & Jefferson, T. Vaccines for preventing tick-borne encephalitis. *Cochrane Database Syst. Rev.* CD000977 (2000).
30. Erra, E. O. & Kantele, A. The Vero cell-derived, inactivated, SA14-14-2 strain-based vaccine (Ixiaro) for prevention of Japanese encephalitis. *Expert Rev. Vaccines* **14,** 1167–1179 (2015).

Acknowledgements We thank P. Vasconcelos, K. Modjarrad, R. Olson, K. Kabra, C. Kannadka, C. Springer, G. Ballarini, G. Alter, K. Stephenson, F. Stephens, B. Lee, J. Jimenez, A. Chandrashekar, P. Penaloza and C. Cabral for generous advice, assistance and reagents. We acknowledge support from the Ragon Institute of MGH, MIT, and Harvard, the National Institutes of Health (AI095985, AI096040, AI100663, AI124377), and the São Paulo Research Foundation (FAPESP 2011/18703-2, 2014/17766-9). The views expressed in this manuscript are those of the authors and do not represent the official views of the Department of the Army or the Department of Defense.

Author Contributions R.A.L., P.A. and D.H.B. designed the studies. J.P.S.P. and P.M.A.Z. developed the challenge virus. P.A., M.B., D.N., M.K., R.N., N.B.M. and Z.L. produced the DNA vaccines and conducted the virologic assays. R.A.B., R.G.J., K.H.E., N.L.M. and S.J.T. produced the PIV vaccines. R.A.L., M.J.I. and A.B.-Z. conducted the mouse studies. R.A.L., C.A.B., E.T.M., E.N.B., P.B.G., D.J., G.N., J.P.N., L.F.M., R.A.B. and R.G.J. conducted the immunologic assays. D.H.B. wrote the paper with all co-authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.H.B. (dbarouch@bidmc.harvard.edu).

Reviewer Information *Nature* thanks A. Barrett and G. Screaton for their contribution to the peer review of this work.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

Animals. Balb/c, SJL, and C57BL/6 female mice at 6–8 weeks of age were purchased from Jackson Laboratories (Bar Harbour). Mice were vaccinated with 50 μg DNA vaccine in saline without adjuvant by the i.m. route or with 1 μg PIV vaccines with 100 μg alum (Alhydrogel; Brenntag Biosector, Denmark) adjuvant by the i.m. or s.c. routes in a 100 μl volume and were then challenged at week 4 by the i.v. route with 10^5 viral particles (10^2 plaque-forming units (PFU)) ZIKV-BR or ZIKV-PR. Animals were randomly allocated to groups. Immunologic and virologic assays were performed blinded. Sample size was determined to achieve 80% power to detect significant differences in protective efficacy. All animal studies were approved by the BIDMC Institutional Animal Care and Use Committee (IACUC).

DNA vaccines. ZIKV strain BeH815744 (accession number KU365780) was used to design transgenes, which were produced synthetically. Sequences were optimized for enhanced transgene expression. Pre-membrane and envelope (prM-Env; defined as amino acids 216–794 of the polyprotein) or Env alone were cloned into the mammalian expression plasmid pcDNA3.1$^+$ (Invitrogen). Deletion mutants lacked the transmembrane (ΔTM) or stem (Δstem) regions of Env. A Kozak sequence and the Japanese encephalitis virus leader sequence were included[24]. Plasmids were produced with Machery-Nagel endotoxin-free gigaprep kits. Sequences were confirmed by double-stranded sequencing.

PIV vaccine. The ZIKV purified inactivated virus (PIV, also termed ZPIV) vaccine was produced at the Pilot Bioproduction Facility, Walter Reed Army Institute of Research, Silver Spring, MD, USA. The PIV vaccine was based on the Puerto Rican PRVABC59 isolate, which was obtained from the US Centers for Disease Control and Prevention, Fort Collins, CO, USA. The Vero cells used for passage and vaccine production were a derivative of a certified cell line manufactured at The Salk Institute, Swiftwater, PA. After inoculation, virus was collected on days 5 and 7, clarified by centrifugation and depth filter (0.45–0.2 μm), and treated with benzonase. The viral harvest was concentrated with an ultrafilter followed by purification using Captocore chromatography resin. The purified ZIKV was then inactivated with formalin (0.05%) at 22 °C for 7 days. Following inactivation, formalin was removed by dialysis, and the antigen concentration was adjusted. The final PIV vaccine was assessed for infectivity by passage in Vero cells followed by plaque assays to demonstrate inactivation.

ZIKV challenge stocks. ZIKV stocks were provided by University of São Paulo, Brazil (Brazil ZKV2015; ZIKV-BR[11]) and the US Centers for Disease Control and Prevention, USA (Puerto Rico PRVABC59; ZIKV-PR). Both strains were passage number 3. Low-passage-number Vero cells were then infected at a multiplicity of infection (MOI) of 0.01 PFU per cell. Supernatant was screened daily for viral titers and collected at peak growth. Culture supernatants were clarified by centrifugation, and fetal bovine serum was added to 20% final concentration (v/v) and stored at −80 °C. The concentration and infectivity of the stocks were determined by RT–PCR and PFU assays. The viral particle to PFU ratio of both stocks was approximately 1,000.

RT–PCR. *Cap* genes of available ZIKV genomes were aligned using Megalign (DNAstar), and primers and probes to a highly conserved region were designed using primer express v3.0 (Applied Biosystems). Primers were synthesized by Integrated DNA Technologies (Coralville) and probes by Biosearch Technologies (Petaluma). To assess viral loads, RNA was extracted from serum with a QIAcube HT (Qiagen). Reverse transcription and RT–PCR were performed as previously described[18]. The wild-type ZIKV BeH815744 *Cap* gene was used as a standard and was cloned into pcDNA3.1+ vector, and the AmpliCap-Max T7 High Yield Message Maker Kit was used to transcribe RNA (Cellscript). RNA was purified using the RNA clean and concentrator kit (Zymo Research), and RNA quality and concentration was assessed by the BIDMC Molecular Core Facility. Log dilutions of the RNA standard were reverse-transcribed and included with each RT–PCR assay. Viral loads were calculated as virus particles per ml. Assay sensitivity was 100 copies per ml. The infectivity of virus in peripheral blood from ZIKV challenged mice was confirmed by PFU assays.

PFU assay. Vero WHO cells were seeded in a MW6 plate to reach confluency at day 3. Cells were infected with log dilutions of ZIKV for 1 h and overlayed with agar. Cells were stained after 6 days of infection by neutral red staining. Plaques were counted, and titers were calculated by multiplying the number of plaques by the dilution and divided by the infection volume.

Western blot. To assess transgene expression from DNA vaccines, cell lysates obtained 48 h following lipofectamine 2000 (Invitrogen) transient transfection of 293T cells were mixed with reducing sample buffer, heated for 5 min at 100 °C, cooled on ice, and run on a precast 4–15% SDS–PAGE gel (Biorad). Protein was transferred to PVDF membranes using the iBlot dry blotting system (Invitrogen), and the membranes were blocked overnight at 4 °C in PBS-T (Dulbeco's phosphate buffered saline + 0.2% V/V Tween 20 + 5% W/V non-fat milk powder). Following overnight blocking, the membranes were incubated for 1 h with PBS-T containing a 1:5,000 dilution of mouse anti-ZIKV Env monoclonal antibody (BioFront Technologies). Membranes were then washed 3 times with PBS-T and incubated for 1 h with PBS-T containing a 1:1,000 dilution of rabbit anti-mouse horseradish peroxidase (HRP) (Jackson ImmunoResearch). Membranes were then washed 3 times with PBS-T and developed using the Amersham ECL plus western blotting detection system (GE Healthcare).

ELISA. Mouse ZIKV Env ELISA kits (Alpha Diagnostic International) were used to determine endpoint antibody titers using a modified protocol. 96-well plates coated with ZIKV Env protein were first equilibrated at room temperature with 300 μl of kit working wash buffer for 5 min. 6 μl of mouse serum was added to the top row, and threefold serial dilutions were tested in the remaining rows. Samples were incubated at room temperature for 1 h, and plates washed 4 times. 100 μl of anti-mouse IgG HRP-conjugate working solution was then added to each well and incubated for 30 min at room temperature. Plates were washed 5 times, developed for 15 min at room temperature with 100 μl of 3,3′,5,5′–tetramehylbenzidine (TMB) substrate, and stopped by the addition of 100 μl of stop solution. Plates were analysed at 450 nm / 550 nm on a VersaMax microplate reader using Softmax Pro 6.0 software (Molecular Devices). ELISA endpoint titers were defined as the highest reciprocal serum dilution that yielded an absorbance >2-fold over background values.

Neutralization assay. A high-throughput ZIKV microneutralization (MN) assay was developed for measuring ZIKV-specific neutralizing antibodies as a modified version of a qualified dengue virus microneutralization assay used in clinical dengue vaccine trials[17]. Briefly, serum samples were serially diluted threefold in 96-well micro-plates, and 100 μl of ZIKV-PR containing 100 PFU were added to 100 μl of each serum dilution and incubated at 35 °C for 2 h. Supernatants were then transferred to microtiter plates containing confluent Vero cell monolayers (World Health Organization, NICSC-011038011038). After incubation for 4 d, cells were fixed with absolute ethanol: methanol for 1 h at −20 °C and washed three times with PBS. The pan-flavivirus monoclonal antibody 6B6-C1 conjugated to HRP (6B6-C1 was a gift from J. T. Roehrig, CDC) was then added to each well, incubated at 35 °C for 2 h, and washed with PBS. Plates were washed, developed with 3,3′,5,5′ –tetramethylbenzidine (TMB) substrate for 50 min at room temperature, stopped with 1:25 phosphoric acid, and absorbance was read at 450 nm. For a valid assay, the average absorbance at 450 nm of three non-infected control wells had to be ≤ 0.5, and virus-only control wells had to be ≥ 0.9. Normalized absorbance values were calculated, and the MN50 titer was determined by a log mid-point linear regression model. The MN50 titer was calculated as the reciprocal of the serum dilution that neutralized $\geq 50\%$ of ZIKV. Seropositivity was defined as a titer $\geq$ 1:10.

ELISPOT. ZIKV-specific cellular immune responses were assessed by IFNγ ELISPOT assays using pool of overlapping 15-amino-acid peptides covering the prM or Env proteins (JPT). 96-well multiscreen plates (Millipore) were coated overnight with 100 μl per well of 10 μg ml^{-1} anti-mouse IFNγ (BD Biosciences) in endotoxin-free Dulbecco's PBS (D-PBS). The plates were then washed three times with D-PBS containing 0.25% Tween 20 (D-PBS-Tween), blocked for 2 h with D-PBS containing 5% FBS at 37 °C, washed three times with D-PBS-Tween, rinsed with RPMI 1640 containing 10% FBS to remove the Tween 20, and incubated with 2 μg ml^{-1} of each peptide and 5×10^5 mouse splenocytes in triplicate in 100 μl reaction mixture volumes. Following 18 h incubation at 37 °C, the plates were washed nine times with PBS-Tween and once with distilled water. The plates were then incubated with 2 μg ml^{-1} biotinylated anti-mouse IFNγ (BD Biosciences) for 2 h at room temperature, washed six times with PBS-Tween, and incubated for 2 h with a 1:500 dilution of streptavidin-alkaline phosphatase (Southern Biotechnology Associates). Following five washes with PBS-Tween and one with PBS, the plates were developed with nitroblue tetrazolium-5-bromo-4-chloro-3-indolyl-phosphate chromogen (Pierce), stopped by washing with tap water, air dried, and read using an ELISPOT reader (Cellular Technology Ltd). The numbers of spot-forming cells (SFC) per 10^6 cells were calculated. The medium background levels were typically <15 SFC per 10^6 cells.

Intracellular cytokine staining. ZIKV-specific CD4$^+$ and CD8$^+$ T-lymphocyte responses were assessed using splenocytes and analysed by flow cytometry. Cells were stimulated for 1 h at 37 °C with 2 μg ml^{-1} of overlapping 15-amino-acid peptides covering the prM or Env proteins (JPT). Following incubation, brefeldin-A and monensin (BioLegend) were added, and samples were incubated for 6 h at 37 °C. Cells were then washed, stained, permeabilized with Cytofix/Cytoperm (BD Biosciences). Data was acquired using an LSR II flow cytometer (BD Biosciences) and analysed using FlowJo v.9.8.3 (Treestar). Monoclonal antibodies included: CD4 (RM4-5), CD8α (53-6.7), CD44 (IM7), and IFNγ (XMG1.2). Antibodies were purchased from BD Biosciences, eBioscience, or BioLegend. Vital dye exclusion (LIVE/DEAD) was purchased from Life Technologies.

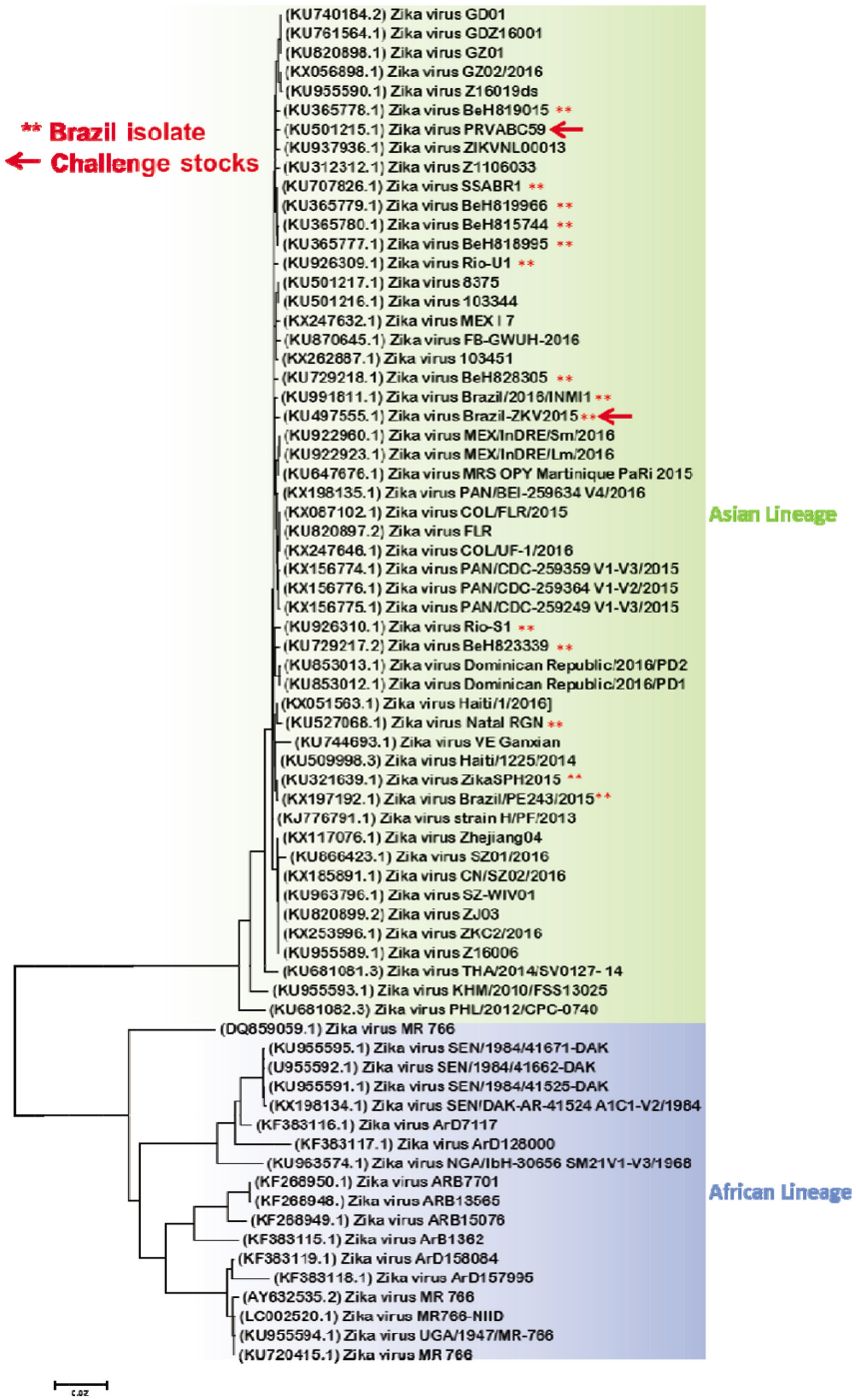
© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

IgG purification and adoptive transfer. Serum was collected from prM-Env DNA-vaccinated mice or naive mice, and polyclonal IgG was purified using protein G purification kits (Thermo Fisher Scientific). Varying amounts of purified IgG was infused by the i.v. route into naive recipient mice before ZIKV challenge.

$CD4^+$ and $CD8^+$ T-lymphocyte depletion. Anti-CD4 (GK1.5) and/or anti-CD8 (2.43) (Bio X Cell) monoclonal antibodies were administered at doses of 500μg per mouse to prM-Env DNA vaccinated mice by the i.p. route on day −2 and day −1 before ZIKV challenge. Antibody depletions were >99.9% efficient as determined by flow cytometry.

Statistical analyses. Analysis of virologic and immunologic data was performed using GraphPad Prism version 6.03 (GraphPad Software). Comparisons of groups was performed using *t*-tests and Wilcoxon rank-sum tests. Correlations were assessed by Spearman rank-correlation tests.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 1 | ZIKV maximum likelihood phylogenetic tree. The ZIKV-BR and ZIKV-PR challenge isolates are depicted with red arrows.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

- Brazil ZKV2015 (Brazil strain; ZIKV-BR challenge stock)
- PRVABC59 (Puerto Rico strain; ZIKV-PR challenge stock)
- BeH815744 (Brazil strain; immunogen design)
- H PF 2013 (French Polynesian strain)
- MR766 (African strain)

AA identical

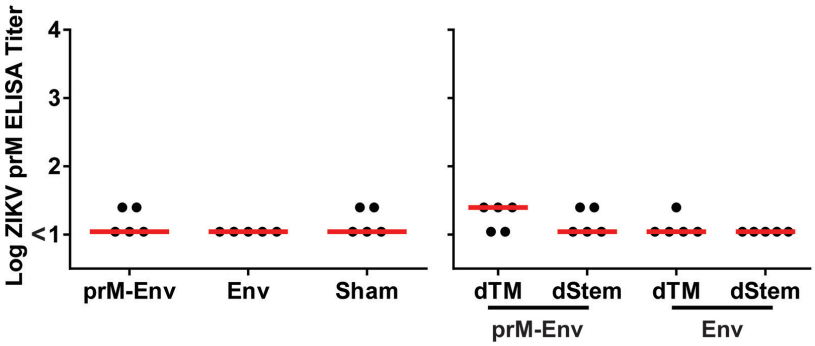
# AA divergent	Brazil ZKV2015	PRVABC59	BeH815744	H PF 2013	MR766
Brazil ZKV2015		3418	3419	3419	3294
PRVABC59	5		3420	3420	3295
BeH815744	4	3		3421	3296
H PF 2013	4	3	2		3298
MR766	125	124	123	121	

Percent identity

Divergence	Brazil ZKV2015	PRVABC59	BeH815744	H PF 2013	MR766
Brazil ZKV2015		99.9	99.9	99.9	96.5
PRVABC59	0.1		99.9	99.9	96.5
BeH815744	0.1	0.1		99.9	96.5
H PF 2013	0.1	0.1	0.1		96.6
MR766	3.6	3.6	3.6	3.5	

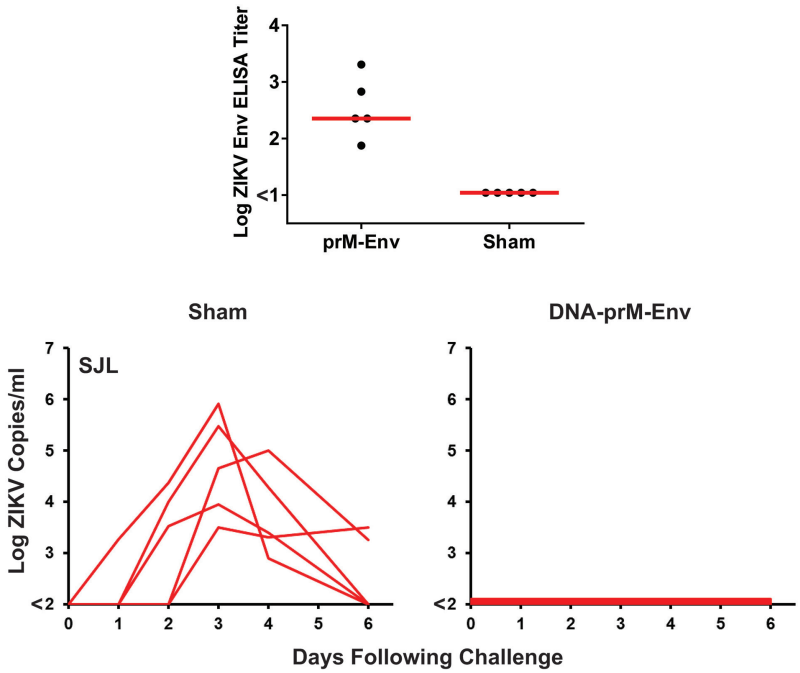
Extended Data Figure 2 | ZIKV amino acid sequence comparisons. Number of and percentage of amino acid differences in the polyprotein are shown for the following ZIKV isolates: Brazil/ZKV2015 (Brazil strain; ZIKV-BR challenge stock), PRVABC59 (Puerto Rico strain; ZIKV-PR challenge stock), BeH815744 (Brazil strain; immunogen design), H/PF/2013 (French Polynesian strain), and MR766 (African strain).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



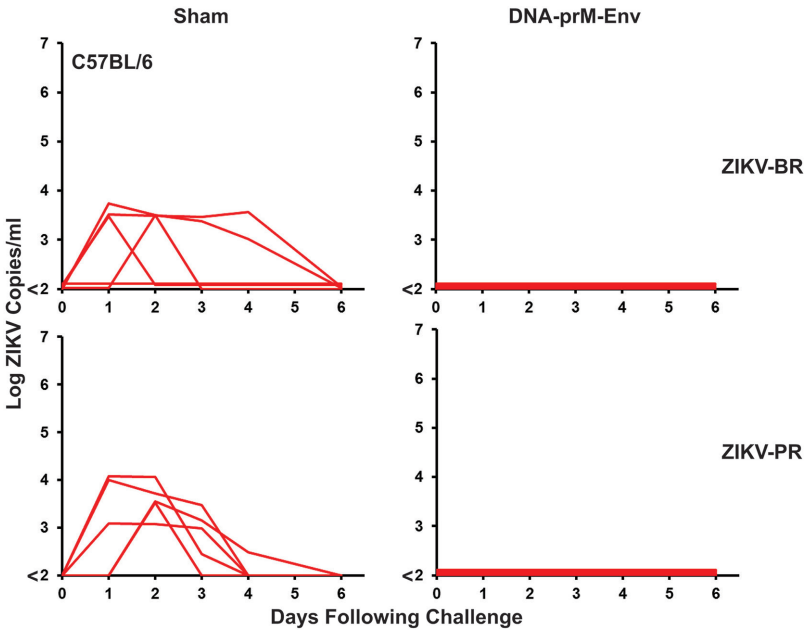
Extended Data Figure 3 | prM-specific antibody responses in DNA-vaccinated mice. In the experiment described in Fig. 2, humoral immune responses were assessed at week 3 following vaccination by prM-specific ELISA. Red bars reflect medians.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



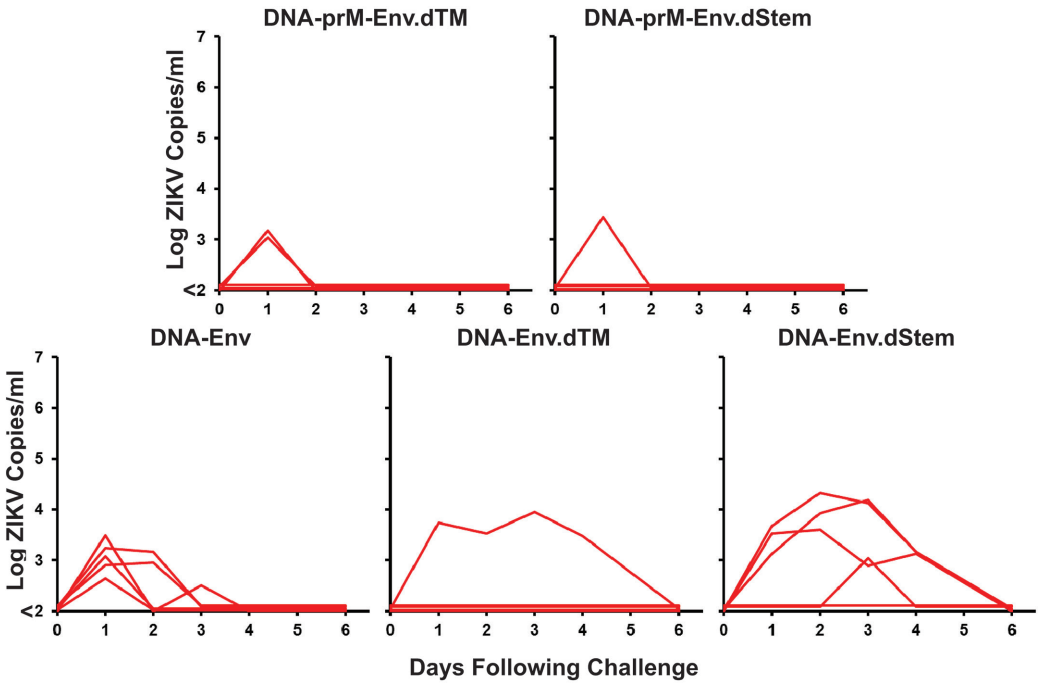
Extended Data Figure 4 | Immunogenicity and protective efficacy of prM-Env DNA vaccine in SJL mice. SJL mice ($n = 5$ per group) received a single immunization by the i.m. route with 50 μg prM-Env DNA vaccine or a sham vaccine and were challenged at week 4 by the i.v. route with 10^5 viral particles (10^2 PFU) ZIKV-BR. Humoral immune responses were assessed at week 3 after vaccination by Env-specific ELISA (top). Red bars reflect medians. Serum viral loads are shown following ZIKV-BR challenge (bottom).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



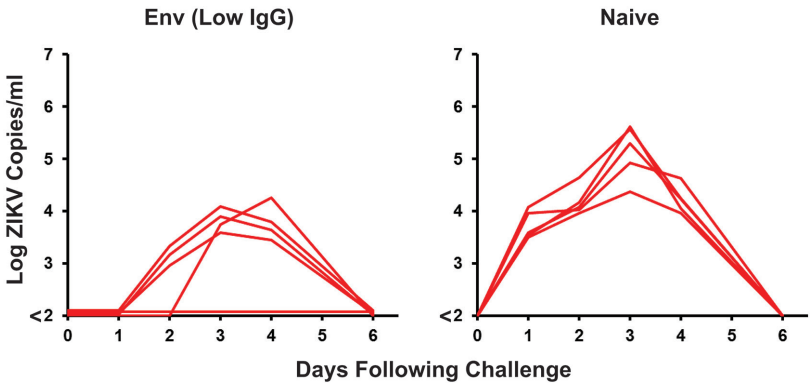
Extended Data Figure 5 | Protective efficacy of prM-Env DNA vaccine in C57BL/6 mice. C57BL/6 mice ($n = 5$ per group) received a single immunization by the i.m. route with 50 μg prM-Env DNA vaccine or a sham vaccine and were challenged at week 4 by the i.v. route with 10^5 viral particles (10^2 PFU) ZIKV-BR or ZIKV-PR. Serum viral loads are shown following challenge.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



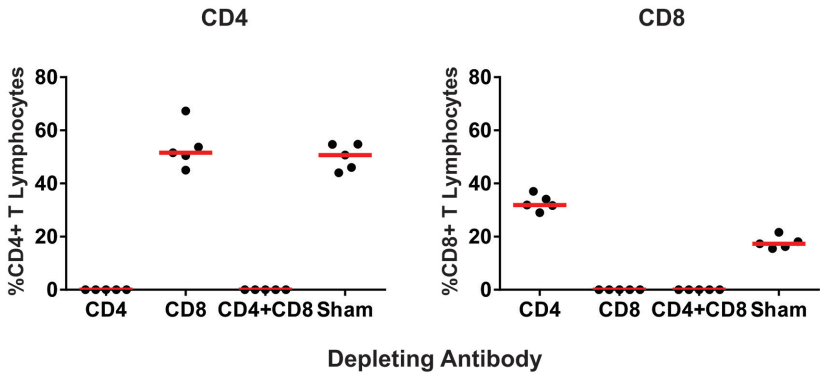
Extended Data Figure 6 | Protective efficacy of various DNA vaccines in C57BL/6 mice. C57BL/6 mice ($n = 5$ per group) received a single immunization by the i.m. route with 50 μg of various DNA vaccines and were challenged at week 4 by the i.v. route with 10^5 viral particles (10^2 PFU) ZIKV-BR. Serum viral loads are shown following challenge.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 7 | Adoptive transfer of low titers of Env-specific IgG. Serum viral loads in mice that received adoptive transfer of low titers of Env-specific IgG (as defined in Fig. 3a) and were then challenged with ZIKV-BR.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 8 | CD4$^+$ and CD8$^+$ T-lymphocyte depletion. CD4$^+$ and/or CD8$^+$ T-lymphocyte depletion following monoclonal antibody treatment of Balb/c mice vaccinated with prM-Env DNA.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

LETTER

doi:10.1038/nature19084

Pancreatic stellate cells support tumour metabolism through autophagic alanine secretion

Cristovão M. Sousa[1], Douglas E. Biancur[1]*, Xiaoxu Wang[1]*, Christopher J. Halbrook[2], Mara H. Sherman[3], Li Zhang[2], Daniel Kremer[2], Rosa F. Hwang[4], Agnes K. Witkiewicz[5,6], Haoqiang Ying[7], John M. Asara[8], Ronald M. Evans[3], Lewis C. Cantley[9], Costas A. Lyssiotis[2,10] & Alec C. Kimmelman[1]†

Pancreatic ductal adenocarcinoma (PDAC) is an aggressive disease characterized by an intense fibrotic stromal response and deregulated metabolism[1–4]. The role of the stroma in PDAC biology is complex and it has been shown to play critical roles that differ depending on the biological context[5–10]. The stromal reaction also impairs the vasculature, leading to a highly hypoxic, nutrient-poor environment[4,11,12]. As such, these tumours must alter how they capture and use nutrients to support their metabolic needs[11,13]. Here we show that stroma-associated pancreatic stellate cells (PSCs) are critical for PDAC metabolism through the secretion of non-essential amino acids (NEAA). Specifically, we uncover a previously undescribed role for alanine, which outcompetes glucose and glutamine-derived carbon in PDAC to fuel the tricarboxylic acid (TCA) cycle, and thus NEAA and lipid biosynthesis. This shift in fuel source decreases the tumour's dependence on glucose and serum-derived nutrients, which are limited in the pancreatic tumour microenvironment[4,11]. Moreover, we demonstrate that alanine secretion by PSCs is dependent on PSC autophagy, a process that is stimulated by cancer cells. Thus, our results demonstrate a novel metabolic interaction between PSCs and cancer cells, in which PSC-derived alanine acts as an alternative carbon source. This finding highlights a previously unappreciated metabolic network within pancreatic tumours in which diverse fuel sources are used to promote growth in an austere tumour microenvironment.

We previously demonstrated that metabolism is rewired in pancreatic cancer cells to facilitate biosynthesis and maintain redox balance in the nutrient-poor conditions of a pancreatic tumour[2,14,15]. While extracellular protein can provide nutrients to the starved cancer cells[11,13], we hypothesized that the stroma may provide additional avenues of metabolic support for the tumour. Pancreatic stellate cells (PSCs) are a predominant cell type in the pancreatic tumour stroma and are important mediators of the desmoplastic response. Their abundance suggests that they may contribute to the metabolism of cancer cells. To test this idea, we assessed changes in the oxygen consumption rate (OCR) and extracellular media acidification rate (ECAR), measures of mitochondrial activity and glycolysis, respectively, in PDAC cells treated with conditioned medium from a well characterized human PSC (hPSC) line[16] (Fig. 1a, b and Extended Data Fig. 1a–e). PDAC glycolysis showed minimal changes when cells were treated with PSC-conditioned medium, as measured by ECAR (Extended Data Fig. 1d, e). By contrast, we observed a consistent increase of 20–40% in the basal OCR after treatment with hPSC medium (Fig. 1a, b and Extended Data Fig. 1a–c), a feature that was independent of serum during the

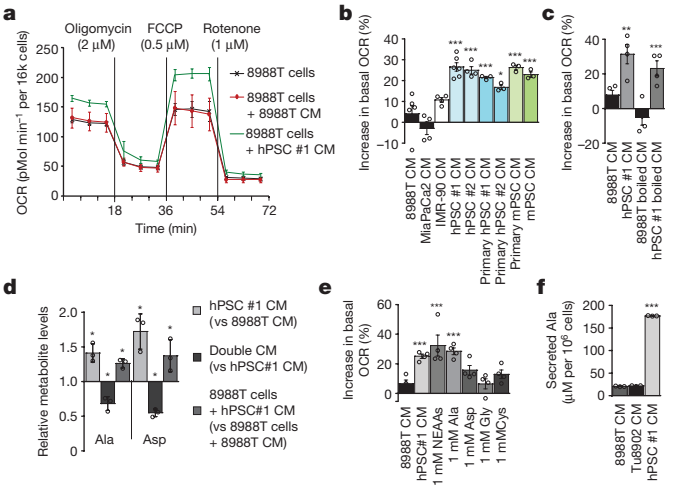


Figure 1 | Pancreatic stellate cells secrete metabolites that fuel pancreatic cancer metabolism. a, Conditioned medium (CM) from hPSCs increases PDAC OCR (green line), as compared to cells treated with PDAC CM (red line) or control (DMEM with 10% serum, black line). A representative trace showing change in OCR during a mitochondrial stress test. Error bars depict s.d. of 6 independent wells from a representative tracing from 6 independent experiments (depicted in **b**). **b**, Per cent change in basal OCR for 8988T cells treated with conditioned medium from different cell lines relative to 8988T cells treated with standard culture medium. Error bars depict s.e.m. of pooled independent experiments ($n = 3$ for primary hPSC #1, #2, primary mPSC; $n = 4$ for hPSC#2, IMR90 and MiaPaCa2; $n = 6$ for 8988T, hPSC#1). **c**, OCR activity of PSC-conditioned medium is retained after heating at 100 °C for 15 min. Error bars, s.e.m. of independent experiments ($n = 4$). **d**, Metabolites that were significantly elevated in PSC-conditioned medium, decreased in double-conditioned medium (PSC-conditioned medium added to 8988T cells and then collected), and elevated intracellularly in PDAC cells treated with PSC-conditioned medium. Error bars, s.d. ($n = 3$). **e**, A mixture of NEAAs (1 mM alanine, aspartate, asparagine, glycine, glutamate, proline and serine) or alanine alone increases PDAC OCR. Data are normalized to cells treated with standard culture medium. Error bars, s.e.m. of independent experiments ($n = 4$). **f**, The concentration of alanine was measured in conditioned medium samples using liquid chromatography with tandem mass spectrometry (LC–MS/MS). Error bars, s.d. ($n = 3$). Significance determined with one-way ANOVA in **b**, **c**, **e**; *t*-test in **d**, **f**. Panels **d**, **f**, $n = 3$ technical replicates from independently prepared samples from individual wells. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. The calculated *P* values and comparisons are reported in Supplementary Information.

[1]Division of Genomic Stability and DNA Repair, Department of Radiation Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02215, USA. [2]Department of Molecular and Integrative Physiology, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA. [3]Gene Expression Laboratory, Howard Hughes Medical Institute, The Salk Institute for Biological Studies, La Jolla, California 92037, USA. [4]Department of Surgical Oncology, The University of Texas M.D. Anderson Cancer Center, Houston, Texas 77230-1402, USA. [5]Department of Pathology, UT Southwestern, Dallas, Texas 75390, USA. [6]Simmons Cancer Center, UT Southwestern, Dallas, Texas 75390, USA. [7]Department of Molecular and Cellular Oncology, UT MD Anderson Cancer Center, Houston, Texas 77030, USA. [8]Division of Signal Transduction, Department of Medicine, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, Massachusetts 02115, USA. [9]Meyer Cancer Center, Department of Medicine, Weill Cornell Medical College, New York, New York 10065, USA. [10]Department of Internal Medicine, Division of Gastroenterology, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA. †Present Address: Department of Radiation Oncology, Perlmutter Cancer Center, NYU Langone Medical Center, New York, New York 10016, USA.
*These authors contributed equally to this work.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

conditioning process (Extended Data Fig. 1f, g) and reproducible with multiple primary specimens (Fig. 1b and Extended Data Fig. 1h, i). Notably, this metabolic phenotype was specific to pancreatic cancer cells; non-transformed pancreatic ductal epithelial cells did not exhibit increased OCR in response to PSC medium (Extended Data Fig. 1j).

To identify the nature of the PSC-secreted factors that alter PDAC metabolism, we subjected conditioned medium to three freeze-thaw cycles (−80 °C, 60 °C) or heating (100 °C, 15 min) and observed that it retained the ability to increase PDAC OCR (Fig. 1c and Extended Data Fig. 1k–m), indicating that the factor(s) lacked tertiary structure. Moreover, the activity was retained in medium passed through a 3-kDa cut-off filter (Extended Data Fig. 1n). The small size and resistance to extreme temperatures of the OCR-increasing factor(s) excluded large candidate molecules such as proteins, indicating that the factor(s) could be metabolites.

We next performed a series of metabolomic studies[17] to identify metabolites that were secreted by PSCs and taken up by PDAC cells (Extended Data Fig. 2a). Specifically, we sought molecules that were over-represented in PSC medium (and therefore secreted by PSCs); under-represented in the PSC medium after contact with PDAC cells (removed by PDAC cells); and over-represented inside PDAC cells treated with the PSC medium (taken up by PDAC cells). Of approximately 200 metabolites analysed, only the non-essential amino acids (NEAA) alanine and aspartate followed this pattern (Fig. 1d and Extended Data Fig. 2b). Treatment of PDAC cells with the individual amino acids revealed that only alanine had the ability to increase PDAC OCR to a degree comparable to that of PSC-conditioned medium (Fig. 1e and Extended Data Fig. 2c).

To demonstrate that PSC-derived metabolites, and alanine specifically, were being taken up by PDAC cells, we performed metabolite tracing experiments in which PSCs were grown to saturation in medium containing uniformly carbon-13-labelled (U^{13}C-)glucose and U^{13}C-glutamine to label secreted NEAAs (Extended Data Fig. 2d). The conditioned medium from the labelled cells was then added to PDAC cells, allowing us to track the production and secretion of alanine by the PSCs and to follow its depletion from PSC medium upon contact with PDAC cells (Extended Data Fig. 2e). Quantification revealed that secreted alanine in the PSC medium reached millimolar concentration within 24 h of conditioning (Fig. 1f and Extended Data Fig. 2f–i), and this occurred independent of serum in the medium (Extended Data Fig. 2h). In the context of a tumour *in vivo*, the more important parameters to assess are the release and uptake of alanine relative to other nutrients. Accordingly, we performed kinetic studies and found that alanine was secreted by the PSCs at the greatest rate of the 14 amino acids measured and more rapidly than even lactate (Extended Data Fig. 2j, k). Alanine was also one of only two amino acids to accumulate in PDAC cells, achieving an enrichment of greater than fivefold (Extended Data Fig. 2l).

We next investigated how PDAC cells metabolize PSC-derived alanine. Alanine could increase OCR by fuelling the TCA cycle, and a likely route for this is through transamination of alanine into pyruvate (Fig. 2a). Consistently, depletion of the cytosolic or mitochondrial alanine transaminase (*GPT1* or *GPT2*, respectively) in PDAC cells (Extended Data Fig. 3a) resulted in an increase in accumulation of alanine (Fig. 2b) and a decrease in OCR in PSC medium-treated PDAC cells (Fig. 2c and Extended Data Fig. 3b, c). Moreover, direct addition of pyruvate to PDAC cells increased OCR (Fig. 2d). We then performed U-^{13}C-Ala tracing studies to assess how alanine was being used in PDAC metabolism. Treatment of cells with 1 mM U-^{13}C-Ala led to a 5–10-fold increase in the intracellular alanine pool (Extended Data Fig. 3d–h). Carbon from alanine did not contribute to upstream glycolytic intermediates (Extended Data Fig. 3i, j) or alter glycolytic flux (Extended Data Fig. 3k) or the NAD$^+$/NADH ratio (Extended Data Fig. 3l, m), and it contributed minimally to the intracellular lactate pool (Extended Data Fig. 3d, n), irrespective of the extracellular glucose concentration (Extended Data Fig. 3o–q). These results suggested

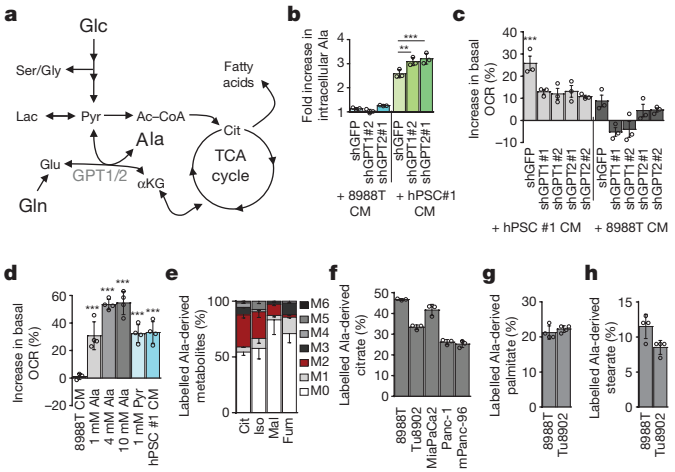


Figure 2 | Alanine is secreted by stellate cells and is used by PDAC to fuel biosynthetic reactions. a, Role of the alanine transaminases (GPT1/2) in central carbon metabolism. Glucose (Glc), pyruvate (Pyr), lactate (Lac), Ac–CoA, citrate (Cit), and α-ketoglutarate (αKG). **b**, Knockdown of *GPT1* or *GPT2* in PDAC cells results in a further increase in intracellular alanine upon treatment with PSC-conditioned medium. Error bars depict s.d. ($n = 3$). Data are normalized to cells treated with standard culture medium for each shRNA. **c**, Knockdown of *GPT1* or *GPT2* in PDAC cells significantly attenuates the ability of PSC-conditioned medium to increase OCR. Data normalized to cells treated with standard culture medium. Error bars depict s.e.m. of independent experiments ($n = 3$). **d**, Alanine treatment of PDAC cells increases OCR in a dose-dependent fashion and can be recapitulated by pyruvate. Data normalized to cells treated with standard culture medium. Error bars depict s.d. of 4 independent wells from a representative experiment (of 3 experiments). **e**, Alanine-derived carbon labelling patterns of metabolites in PDAC cells treated with U-^{13}C-Ala demonstrate substantial label incorporation into the TCA cycle metabolites citrate, isocitrate (Iso), malate (Mal) and fumarate (Fum). Error bars depict s.d. ($n = 3$). **f**, U-^{13}C-alanine labelling of citrate in a panel of PDAC cell lines represented as fraction of citrate with labelled carbon. Error bars depict s.d. ($n = 3$). **g**, **h**, U-^{13}C-Ala labelled PDAC cells show substantial incorporation of alanine into the *de novo* biosynthesis of the fatty acids palmitate (**g**) and stearate (**h**). Data presented as the sum of all isotopomers containing alanine-derived label. Error bars depict s.d. ($n = 4$). Significance determined with one-way ANOVA in **b**–**d**. Panels **b**, **f**, $n = 3$; panels **g**, **h**, $n = 4$ technical replicates from independently prepared samples from individual wells. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. The calculated P values and comparisons are reported in Supplementary Information.

that alanine-derived pyruvate was being used in the mitochondria, because it was not affecting cytosolic glycolytic metabolism. Indeed, alanine was a major carbon source for the TCA cycle; ^{13}C was markedly incorporated into citrate and isocitrate, and, to a lesser extent, malate and fumarate (Fig. 2e and Extended Data Figs 4, 5), as well as into the NEAAs aspartate and glutamate (Extended Data Fig. 4–5), which can be biosynthesized from TCA cycle intermediates in PDAC cells[14]. Indeed, citrate was one of the major recipients of alanine carbon, with labelling ranging from 23% to 46% among PDAC lines (Fig. 2f and Extended Data Figs 4, 5).

We also observed that the alanine-derived pyruvate competed meaningfully with mitochondrial but not cytosolic glucose-derived pyruvate based on the citrate (M2, where M refers to metabolite and 2 the number of ^{13}C atoms present) and lactate (M3) labelling patterns following U-^{13}C-Ala or U-^{13}C-glucose tracing studies (Extended Data Figs 3d, i, j, o–q, 4, 5, 6a). These results illustrate that alanine carbon is being used selectively to fuel mitochondrial metabolism and are consistent with our earlier observations that the addition of alanine does not disrupt glycolysis (Extended Data Fig. 3i–k). A principle function of mitochondrial pyruvate in proliferating cells is to contribute to citrate generation via its conversion to acetyl coenzyme A

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

(Ac-CoA). This citrate (labelled with the two Ac-CoA carbons, M2) is then released into the cytosol and used as a building block for fatty acid biosynthesis. As shown in Fig. 2g, h, more than 20% of total palmitate and 10% of total stearate contained alanine-derived ^{13}C, rivalling the contribution derived from glucose (around 50% and 20%, respectively; Extended Data Fig. 6b–g), and illustrating the importance of alanine in lipid biosynthesis. The production of citrate from alanine for fatty acid biosynthesis requires an additional TCA cycle input to support anaplerosis. Indeed, we observed that alanine could also serve in this role through pyruvate carboxylase, as evidenced by M3 labelling of citrate and other TCA cycle intermediates from U-^{13}C–Ala (Extended Data Figs 4, 5).

We also found that addition of alanine increased flux through the serine biosynthetic pathway in PDAC cells (Extended Data Fig. 6h–j). This pathway uses glucose to generate the NEAAs serine and glycine, which are used for, among other processes, the *de novo* biosynthesis of nucleic acids. U-^{13}C-glucose tracing studies confirmed that alanine promoted flux of glucose-derived carbon through this pathway (Extended Data Fig. 6h), an effect that was more pronounced under the glucose-limiting conditions, which resemble poorly perfused pancreatic tumours (Extended Data Fig. 6j). Collectively, these results demonstrated that alanine-derived carbon could supplant glucose-derived carbon in TCA cycle metabolism (to make lipids and NEAAs), thereby enabling glucose to be used for additional biosynthetic functions such as the serine biosynthetic pathway. Glutamine metabolism was also altered in the presence of exogenous alanine, with more alanine being incorporated into the TCA cycle in place of glutamine carbon (Extended Data Fig. 6k, l), but this occurred to a much lower degree relative to glucose-derived carbon in most PDAC lines.

One possible source of PSC-secreted alanine was protein catabolism. Given the critical role of autophagy in PDAC[18,19], we tested whether the secretion of alanine was autophagy-dependent. PSCs in culture exhibit readily detectable levels of basal autophagy, as demonstrated by LC3 puncta (representing autophagosomes; Extended Data Fig. 7a), and exhibit appreciable levels of autophagic flux (Fig. 3a–c and Extended Data Fig. 7b–d). Notably, co-culture with PDAC cells (Fig. 3a, b), or treatment with PDAC-conditioned medium (Fig. 3c and Extended Data Fig. 7d), significantly increased autophagic flux in PSCs. To assess the importance of autophagy in the secretion of alanine, we inhibited autophagy in PSCs using short hairpin (sh)RNAs targeting the essential autophagy genes *ATG5* and *ATG7* (Extended Data Fig. 7b, c, e, f) and measured the ability of the PSC-conditioned medium to increase PDAC OCR. The ability of PSC medium to increase OCR in PDAC cells was abolished upon knockdown of autophagy genes (Fig. 3d and Extended Data Fig. 7g) and could be rescued by addition of exogenous alanine (Extended Data Fig. 7h). Consistently, we observed a drop in the levels of both intracellular alanine and secreted alanine when autophagy was impaired in hPSCs (Fig. 3e and Extended Data Fig. 7i–m). Furthermore, intracellular alanine accumulation in PDAC cells was also diminished following treatment with autophagy-impaired PSC-conditioned medium (Fig. 3f). In contrast to PDAC cells, which respond robustly to autophagy inhibition[18,19], PSC proliferation was largely insensitive to autophagy impairment (Extended Data Fig. 7n). Even under nutrient-depleted conditions, PSC survival was not altered when autophagy was impaired (Extended Data Fig. 7o, p), indicating that altered PSC survival was not the cause of decreased alanine secretion. In sum, these data provide evidence for two-way intratumoral metabolic cross-talk, in which cancer cells release a signal to PSCs that results in the induction of autophagy and the release of alanine.

We next sought to determine whether these PSC-induced metabolic alterations in PDAC would have an impact on PDAC proliferation. Not surprisingly, in high-nutrient medium (25 mM glucose, 4 mM Gln, full serum), the impact of PSC medium on PDAC proliferation was not significant (Extended Data Fig. 8a). By contrast, when grown in a low-nutrient setting (serum-free or glucose-limiting conditions), recapitulating the austere

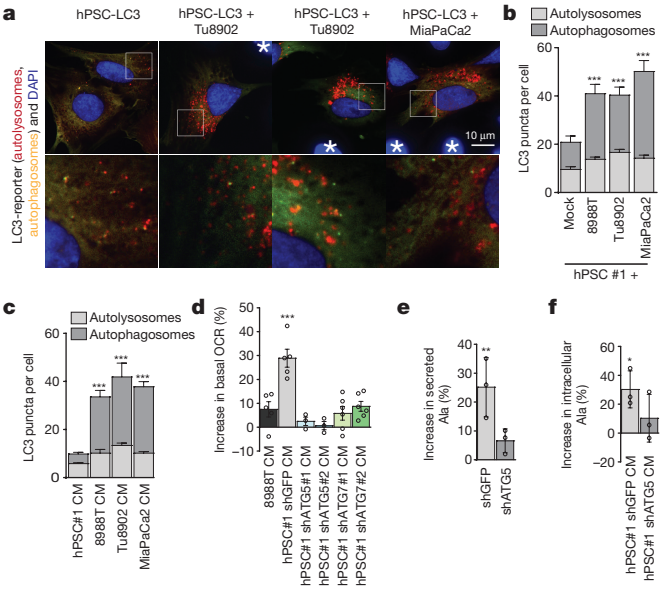


Figure 3 | Alanine secretion is dependent on stellate cell autophagy. **a**, Representative images of basal autophagic flux in PSCs using an LC3 tandem fluorescence (GFP–RFP) reporter. PDAC cells are not fluorescently labelled and are indicated by asterisks. **b**, Autophagy in PSCs is increased upon co-culture with multiple PDAC lines, demonstrated by increased autolysosomes. Error bars, s.e.m. of $n = 32$ for hPSC alone; $n = 38$ for 8988T co-culture; $n = 41$ for Tu8902 co-culture; $n = 24$ for MiaPaCa2 co-culture pooled from 3 independent experiments. **c**, Quantification of autophagic flux in stellate cells treated with conditioned medium from PDAC lines. Error bars, s.e.m. of $n = 65$ hPSC-conditioned medium; $n = 50$ 8988T-conditioned medium; $n = 61$ Tu8902-conditioned medium; $n = 62$ MiaPaca2-conditioned medium in independent hPSC cells per condition pooled from 3 independent experiments. **d**, Suppression of PSC autophagy by *ATG5* or *ATG7* knockdown attenuates the ability of PSC-conditioned medium to increase OCR in 8988T PDAC cells. Data normalized to cells treated with standard culture medium. Error bars show s.e.m. of pooled experiments ($n = 6$ for ATG7 #1, #2; $n = 5$ for shGFP, 8988T conditioned medium; $n = 3$ for ATG5 #1, #2). **e**, PSC-conditioned medium contains elevated alanine, which is significantly decreased when autophagy is inhibited. **f**, PDAC cells treated with PSC conditioned medium display elevated intracellular alanine, which is significantly suppressed when autophagy is inhibited in PSCs. Error bars for **e**, **f**, show s.d. of $n = 3$ technical replicates from independently prepared samples from individual wells. Significance determined with two-way ANOVA in **b**, **c**; one-way ANOVA in **d**–**f**. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. The calculated P values and comparisons are reported in Supplementary Information.

microenvironment *in vivo*, there was a significant positive effect of PSC medium on PDAC proliferation (Fig. 4a and Extended Data Fig. 8b–f). Notably, PSC medium from autophagy-impaired cells did not produce this growth-promoting effect (Fig. 4a and Extended Data Fig. 8d–i). Supplementation of serum-free or low-glucose medium with alanine was also able to rescue growth in a manner similar to PSC medium (Fig. 4a and Extended Data Fig. 8d–f, j–l), while having a minimal effect in nutrient-rich settings (Extended Data Fig. 8a). Moreover, alanine was able to restore the growth-promoting ability of PSC medium from autophagy-impaired cells (Extended Data Fig. 8g). Consistent with the proposed mechanism, pyruvate (the product of the transamination of alanine), but not lactate, was able to sustain PDAC proliferation in glucose-limited medium (Extended Data Fig. 8j–l). Similarly, *GPT1* knockdown in PDAC cells grown under serum-free conditions significantly attenuated the effects of PSC medium or alanine on proliferation (Extended Data Fig. 8m).

To determine whether this cross-feeding mechanism was operative *in vivo*, we developed a co-injection system in which PSCs could be manipulated genetically and then implanted alongside cancer cells

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

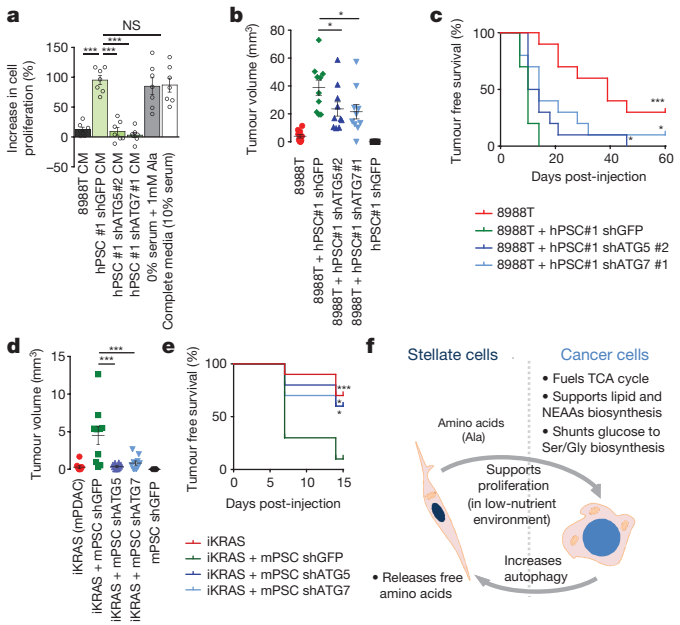


Figure 4 | Stellate cell metabolite secretion supports PDAC growth under nutrient-limiting conditions and facilitates tumour growth. **a**, PSC-conditioned medium increases proliferation of 8988T cells grown in serum-free medium, a feature recapitulated with exogenous alanine supplementation. By contrast, conditioned medium from PSCs where autophagy is inhibited is unable to support PDAC growth. The addition of 10% serum is included as a positive control. Error bars, s.e.m. ($n = 7$) independent experiments. **b**, **c**, Co-injection of 8988T PDAC cells with PSCs significantly enhances early tumour growth, analysed at 35 days post-injection (**b**), and decreases tumour-free survival (**c**), in a xenograft model. This effect is significantly attenuated when autophagy is suppressed in the PSCs. Error bars, s.e.m. ($n = 10$) tumours per condition per time point, except for the PSC–shGFP control, for which ($n = 5$) animals were injected. **d**, **e** Co-injection of inducible KRAS mouse PDAC cells with mPSCs in the pancreas of syngeneic mice significantly enhances early tumour growth at 7 days post-injection (**d**) and decreases tumour-free survival (**e**) in an orthotopic syngeneic graft model, whereas this effect is significantly attenuated when autophagy is suppressed in the PSCs. Error bars, s.e.m. ($n = 10$) tumours per condition per time-point, except for the mPSC–shGFP control, for which ($n = 5$) animals were injected. **f**, Model of tumour–stroma metabolic cross-talk. Significance determined with a one-way ANOVA in **a**, **b**, **d**; log-rank Mantel–Cox test in **c**, **e**. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. The calculated P values and comparisons are reported in Supplementary Information.

into the flanks of mice. Although previous studies have shown that co-injection of PSCs with PDAC cells promotes tumour growth[16], the contribution of metabolic cross-talk to this effect has not been explored. The fact that autophagy is required for PSC alanine secretion, while having a minimal impact on proliferation and no effect on survival (Extended Data Fig. 7n–p), permits selective attenuation of this metabolic cross-talk so that its role in tumour growth can be assessed. We therefore performed co-injection studies with limiting numbers of PDAC cells along with autophagy-competent or incompetent PSCs (Extended Data Fig. 9a). Consistent with the *in vitro* proliferation data, both tumour take and tumour growth kinetics were increased significantly when co-injected with autophagy-competent PSCs, and this increase was significantly attenuated when PDAC cells were co-injected with autophagy-incompetent PSCs (Fig. 4b, c and Extended Data Fig. 9b–e). Notably, the predominant effect of PSC growth support in these assays appears to be during tumour initiation or tumour take (Fig. 4b, c and Extended Data Fig. 9c, d); at later time points the autophagy-incompetent PSC-containing tumours approximate their autophagy-competent counterparts (Extended Data Fig. 9b, e). Indeed, the cancer cells ultimately form the majority of the grafted tumours, even with the wild-type PSCs (Extended Data Fig. 9f, g). Quantification of stellate cells by western blot for RFP (exogenous PSCs were marked with RFP) showed that while the levels of RFP expression varied by individual tumour, there were no significant differences between groups based on autophagy status (Extended Data Fig. 9h, i).

Consistent with the results from the subcutaneous model, diminution of autophagy in PSCs slowed tumour growth and increased tumour-free survival in orthotopic assays (Extended Data Fig. 10a–d). The orthotopic tumours contained more PSCs than the subcutaneous tumours, but again PSCs were less abundant than tumour cells (Extended Data Fig. 10e). To further validate the *in vivo* relevance of this mechanism, we used a syngeneic system consisting of cancer cells and PSCs derived from our murine PDAC autochthonous models[15]. Consistent with our findings using human PDAC and PSC lines, murine PSCs improved tumour engraftment in an autophagy-dependent manner (Fig. 4d, e and Extended Data Fig. 7l, m).

Our work demonstrates growth-promoting metabolic cross-talk between stromal PSCs and epithelial cancer cells (Fig. 4f). Previous work has illustrated that intratumoral metabolic cross-talk can occur between different populations of cells in a tumour[20,21]. For example, hypoxic cancer cells in colorectal and lung cancer models preferentially consume glucose and provide well-oxygenated cancer cells with lactate for oxidative metabolism[21]. More recently, others have suggested that stromal cells can alter the metabolism of pancreatic cancer cells[22] and that this can occur with exosomes as a vehicle for non-selective metabolite delivery[20]. While alterations in lactate and alanine were recently reported using imaging studies during pancreatic cancer progression in mouse models, these studies did not have the spatial resolution to distinguish between different cell types within the tumour and would not be able to detect metabolic cross-talk between cancer cells and fibroblasts[23]. Our work, using pancreatic cancer model systems and *in vivo* co-transplantation assays, demonstrate a specific role of PSC-derived alanine, whose carbon competes with glucose-derived carbon as a metabolic substrate—effects not recapitulated with exogenous lactate.

Protein catabolized through autophagy is the source of secreted alanine from PSCs. Given that alanine is the second most represented amino acid in proteins, catabolism can provide an important source of free alanine[24]. We found that PDAC cells stimulate autophagy in PSCs and selectively consume the released alanine, using this carbon in the TCA cycle following its conversion to pyruvate. Notably, alanine-derived pyruvate does not equilibrate with cytosolic pyruvate. As such, it provides a direct mitochondrial carbon source without changing the cytosolic NAD^+/NADH redox balance. This results in increased mitochondrial oxygen consumption as well as lipid biosynthesis. Importantly, the metabolism of alanine allows traditional carbon sources such as glucose to be more available for other biosynthetic processes, such as the synthesis of serine and glycine, which are precursors for nucleic acid biosynthesis. This cooperative metabolism promotes tumour growth and is particularly relevant in nutrient-limited conditions.

According to this mechanism, we would predict that PSC-mediated alanine secretion would be less relevant under circumstances in which pancreatic cancer cells are able to access nutrients through the vasculature directly, such as those in PSC-ablated pancreatic tumours[8,9]. In contrast, alteration of PSC features (inhibiting autophagy, for example) without their ablation would allow the continued, PSC-mediated restraint of tumour growth while potentially sensitizing the tumour to treatment. A recent example of this strategy demonstrated that stromal reprogramming of PDAC with vitamin D sensitized them to chemotherapy[25].

Collectively, these findings are important for the following reasons. First, they shed light on the complex biology of tumour–stroma interactions. Furthermore, they highlight the importance of studying the metabolism of highly desmoplastic tumours such as PDAC in the proper context, in this case in the presence of relevant supporting cell types. Last, this cooperative metabolism between cancer cells and PSCs,

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

which is dependent on autophagy, further substantiates the critical role of metabolic scavenging in PDAC and suggests that inhibition of autophagy and other lysosomal degradation pathways may have an even greater therapeutic utility in this disease than previously appreciated.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 9 October 2015; accepted 7 July 2016.
Published online 10 August 2016.

1. Hidalgo, M. Pancreatic cancer. *N. Engl. J. Med.* **362,** 1605–1617 (2010).
2. Sousa, C. M. & Kimmelman, A. C. The complex landscape of pancreatic cancer metabolism. *Carcinogenesis* **35,** 1441–1450 (2014).
3. Whatcott, C. J. *et al.* Desmoplasia in primary tumors and metastatic lesions of pancreatic cancer. *Clin. Cancer Res.* **21,** 3561–3568 (2015).
4. Feig, C. *et al.* The pancreas cancer microenvironment. *Clin. Cancer Res.* **18,** 4266–4276 (2012).
5. Erkan, M. *et al.* StellaTUM: current consensus and discussion on pancreatic stellate cell research. *Gut* **61,** 172–178 (2012).
6. Gore, J. & Korc, M. Pancreatic cancer stroma: friend or foe? *Cancer Cell* **25,** 711–712 (2014).
7. Apte, M. V & Wilson, J. S. Dangerous liaisons: pancreatic stellate cells and pancreatic cancer cells. *J. Gastroenterol. Hepatol.* **27** (Suppl 2), 69–74 (2012).
8. Özdemir, B. C. *et al.* Depletion of carcinoma-associated fibroblasts and fibrosis induces immunosuppression and accelerates pancreas cancer with reduced survival. *Cancer Cell* **25,** 719–734 (2014).
9. Rhim, A. D. *et al.* Stromal elements act to restrain, rather than support, pancreatic ductal adenocarcinoma. *Cancer Cell* **25,** 735–747 (2014).
10. Feig, C. *et al.* Targeting CXCL12 from FAP-expressing carcinoma-associated fibroblasts synergizes with anti-PD-L1 immunotherapy in pancreatic cancer. *Proc. Natl Acad. Sci. USA* **110,** 20212–20217 (2013).
11. Kamphorst, J. J. *et al.* Human pancreatic cancer tumors are nutrient poor and tumor cells actively scavenge extracellular protein. *Cancer Res.* **75,** 544–553 (2015).
12. Olive, K. P. *et al.* Inhibition of Hedgehog signaling enhances delivery of chemotherapy in a mouse model of pancreatic cancer. *Science* **324,** 1457–1461 (2009).
13. Commisso, C. *et al.* Macropinocytosis of protein is an amino acid supply route in Ras-transformed cells. *Nature* **497,** 633–637 (2013).
14. Son, J. *et al.* Glutamine supports pancreatic cancer growth through a KRAS-regulated metabolic pathway. *Nature* **496,** 101–105 (2013).
15. Ying, H. *et al.* Oncogenic Kras maintains pancreatic tumors through regulation of anabolic glucose metabolism. *Cell* **149,** 656–670 (2012).
16. Hwang, R. F. *et al.* Cancer-associated stromal fibroblasts promote pancreatic tumor progression. *Cancer Res.* **68,** 918–926 (2008).
17. Yuan, M., Breitkopf, S. B., Yang, X. & Asara, J. M. A positive/negative ion–switching, targeted mass spectrometry–based metabolomics platform for bodily fluids, cells, and fresh and fixed tissue. *Nat. Protocols* **7,** 872–881 (2012).
18. Yang, A. *et al.* Autophagy is critical for pancreatic tumor growth and progression in tumors with p53 alterations. *Cancer Discov.* **4,** 905–913 (2014).
19. Yang, S. *et al.* Pancreatic cancers require autophagy for tumor growth. *Genes Dev.* **25,** 717–729 (2011).
20. Zhao, H. *et al.* Tumor microenvironment derived exosomes pleiotropically modulate cancer cell metabolism. *eLife* **5,** e10250 (2016).
21. Sonveaux, P. *et al.* Targeting lactate-fueled respiration selectively kills hypoxic tumor cells in mice. *J. Clin. Invest.* **118,** 3930–3942 (2008).
22. Tape, C. J. *et al.* Oncogenic KRAS regulates tumor cell signaling via stromal reciprocation. *Cell* **165,** 910–920 (2016).
23. Serrao, E. M. *et al.* MRI with hyperpolarised [1-^{13}C] pyruvate detects advanced pancreatic preneoplasia prior to invasive disease in a mouse model. Gut http://dx.doi.org/10.1136/gutjnl-2015–310114 (2015).
24. Trinquier, G. & Sanejouand, Y. H. Which effective property of amino acids is best preserved by the genetic code? *Protein Eng.* **11,** 153–169 (1998).
25. Sherman, M. H. *et al.* Vitamin D receptor-mediated stromal reprogramming suppresses pancreatitis and enhances pancreatic cancer therapy. *Cell* **159,** 80–93 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank Kimmelman laboratory members for reading the manuscript, A. Yang for help with orthotopic injections, M. Yuan and S. Breitkopf for technical support with mass spectrometry, E. Sicinska for resected PDAC specimens, and the Dana-Farber/Harvard Cancer Center Rodent Histopathology Core for assistance with tissue processing. DFHCC is supported in part by NIH 5P30CA06516. A.C.K. is supported by NIH grant GM095567, NCI grants R01CA157490, R01CA188048, ACS Research Scholar Grant (RSG-13-298-01-TBG), and the Lustgarten Foundation. C.A.L. is supported by a PanCAN-AACR Pathway to Leadership award and a Dale F. Frey award from the Damon Runyon Cancer Research Foundation (DFS-09-14). Metabolomics studies performed at the University of Michigan were supported by NIH grant DK097153. L.C.C. was supported by P01CA117969, PanCAN-AACR and the Lustgarten Foundation. J.M.A. was supported by P30CA006516 and P01CA120964. R.M.E. is an investigator of the Howard Hughes Medical Institute and March of Dimes Chair in Molecular and Developmental Biology at the Salk Institute, and is supported in part by grants from The Lustgarten Foundation and a Stand Up to Cancer Dream Team Translational Cancer Research Grant, a Program of the Entertainment Industry Foundation (SU2C-AACR-DT0509).

Author Contributions C.M.S. performed or participated in the conception and performance of all experiments. D.E.B. and X.W. participated in tissue culture imaging and animal studies. C.J.H. conducted kinetic metabolite release, capture and quantification metabolomics studies with support from D.K. L.Z. performed the associated metabolomics and fatty acid analyses. J.M.A. supervised the metabolite tracing metabolomics experiments as a core service. M.H.S., D.K., R.F.H, A.K.W. and H.Y. provided essential reagents. R.M.E. and L.C.C. provided intellectual feedback and support. C.M.S., C.A.L. and A.C.K. conceived the project, planned and guided the research, and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.C.K. (alec.kimmelman@nyumc.org) or C.A.L. (clyssiot@med.umich.edu).

Reviewer Information *Nature* thanks J. Debnath and the other anonymous reviewer(s) for their contribution to the peer review of this work.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

Cell culture. The cell lines 8988T, MiaPaCa2, Tu8902, Panc1, MPanc96 and IMR90 were obtained from ATCC or the DSMZ. hPSCs (hPSC#1) have been previously described[16]. hPSC#2 was isolated from an untreated human PDAC resection and considered de-identified 'surgical waste' tissue under IRB approved protocols 03-189 and 11-104. Patients gave informed consent for tissue collection. Stromal cells that outgrew the cancer cells in culture were isolated by differential trypsinization and immortalized by infection with hTERT and SV40gp6 (Addgene plasmids #22396 and #10891, respectively) retro-viruses. These cells were kept in DMEM (Life Technologies 11965) supplemented with 10% FBS and 1% Pen/Strep (Life Technologies 15140). Primary human pancreatic cancer-associated fibroblasts were isolated from tumour resections in a similar manner as above, under IRB approved protocol STU 102010-051, but were not immortalized. Cells were kept in DMEM supplemented with 10% CCS (Thermo Scientific) and 1% Pen/Strep. PSCs were verified by measuring Desmin and SMA expression. HPDE have been previously described and grown as indicated[26]. All cells were routinely tested for mycoplasma by PCR and PDAC lines were typically authenticated by fingerprinting as well as visual inspection and carefully maintained in a centralized cell bank. mPSCs were isolated from normal mice pancreata and purified by centrifugation using a Nycodenz gradient and activated by *in vitro* culture, as described[25].

Black 6 (B6) mPSCs were generated from B6 females (Taconic, B6NTac) harbouring mouse PDAC. These animals were pre-treated with doxycycline diet and kept in doxycycline regimen for the duration of the experiment and were injected with 5×10^5 iKRAS mPDAC cells[15] into the pancreas. Pancreatic tumours were resected at 2 weeks, digested in collagenase and dispase and mechanically minced. Cells were plated in cell culture dishes in DMEM (Gibco) with 15% FBS in the absence of doxycycline to limit the growth of iKRAS mouse PDAC cells. mPSCs were immortalized by infection with hTERT and SV40gp6 (Adgene plasmids #22396 and #10891, respectively) retro-viruses. Cells were kept in DMEM (Gibco) with 10% FBS and 1% Pen/strep.

mCherry-hPSCs are hPSC#1 labelled with mCherry through infection with a lentivirus expressing mCherry.

Conditioned medium was generated by adding fresh medium to cells at >50% confluence. Medium was harvested 48 h later and passed through 0.45-μm filters. For size cut-off experiments conditioned medium was filtered through 3-kDa cutoff columns (EMD Millipore, UFC900308). Concentrated (>3 kDa) medium was resuspended in a DMEM volume matching the initial medium volume. Boiled medium experiments were performed by heating conditioned medium at 100 °C for 15 min followed by filtration at 0.45 μm to remove precipitate. Freeze-thaw medium was treated by 3 consecutive cycles of 15 min at −80 °C followed by 15 min at 60 °C and then filtered to remove precipitate.

The tandem fluorescence LC3-reporter stable hPSC cells were generated by retroviral infection of hPSCs with pBABE-puro mCherry-EGFP-LC3B (Addgene plasmid #22418). For autophagic flux quantification experiments, 7.5×10^4 hPSC-LC3 cells were plated in 12-well plates with cover slips and 3×10^5 PDAC or hPSC cells were added 4 h later. Cover slips were fixed in 4% paraformaldehyde (ThermoFisher, 28908) after cells had been in contact for 24 h. Coverslips were mounted in DAPI containing mounting solution (Life Technologies P36931). Cells were imaged on a Yokogawa Spinning Disk Confocal in FITC, RFP and DAPI channels. The ratio of red:yellow puncta was determined by counting puncta using the Cell Counter imageJ plugin.

Oil-red O staining was performed on cells plated on glass cover slips and fixed 24 h after plating in 4% paraformaldehyde (Thermo-Fisher, 28908) for 15 min. Cells were rinsed with PBS followed by a rinse with 60% isopropanol and stained with freshly prepared Oil Red O working solution comprised of 3 ml of 0.5% solution (Sigma, O1391) and 2 ml of H_2O for 15 min, rinsed with 60% isopropanol and counterstained with Heamatoxylin. Cover slips were then washed in H_2O, mounted in Vectashield and imaged using a Leica DM2000 bright-field microscope.

Growth assays. Growth curves were obtained as previously described[19]. Cell growth over 48 h was assessed in clear bottom 96-well plates (Costar 3603, Corning Incorporated) by CellTiter-Glo (Promega G7572) analysis 48 h post treatment with conditioned medium or metabolites and determined by the mean of at least three wells per condition. Luminescence was measured on a POLARstar Omega plate reader.

Metabolism experiments. OCR and ECAR experiments were performed using the XF-96 apparatus from Seahorse Bioscience. Cells were plated (16,000 cells per well for 8988T; 20,000 cells for Tu8902 or Panc-1; 50,000 cells for HPDE) in at least quadruplicate for each condition the day before the experiment. The next day, medium was completely replaced with conditioned medium (75 μl of conditioned medium and 25 μl of fresh medium) or fresh medium containing either 1 mM L-alanine (Sigma A7469), 1 mM of NEAAs (Gibco 11140), 1 mM glycine (Sigma G8790), 1 mM aspartate (Sigma A4534) or 1 mM cysteine (Sigma A9165). 20 h later, medium was replaced by reconstituted DMEM with 25 mM glucose and 2 mM glutamine (no sodium bicarbonate) adjusted to pH∼7.4 and incubated for 30 min at 37% in a CO_2-free incubator. For the mitochondrial stress test (Seahorse 101706-100), oligomycin, FCCP and rotenone were injected to a final concentration of 2 μM, 0.5 μM and 4 μM, respectively. For the glycolysis stress test (Seahorse 102194-100), glucose, oligomycin and 2-deoxyglucose were injected to a final concentration of 10 mM, 2 μM and 100 mM, respectively. OCR and ECAR were normalized to cell number as determined by CellTiter-Glo analysis at the end of the experiments.

Metabolomics. Steady-state metabolomics experiments were performed as previously described[14]. Briefly, PDAC cell lines were grown to ∼80% confluence in growth medium (DMEM, 2 mM glutamine, 10 mM glucose, 10% CCS) on 6 cm dishes in biological triplicate. A complete medium change was performed two hours before metabolite collection. To trace the effect of alanine on glutamine and glucose metabolism, PDAC cell lines were grown as above and then transferred into glutamine-free (with 10 mM glucose) or glucose-free (with 2 mM glutamine) DMEM containing 10% dialysed FBS and supplemented with either 2 mM U-^{13}C-glutamine (± 1 mM alanine) or 10 mM U-^{13}C-glucose (± 1 mM alanine), respectively, overnight. To trace alanine metabolism, PDAC cell lines were grown as above and then transferred into DMEM (with 10 mM glucose, 2 mM glutamine, 10% dialysed FBS) and supplemented with 1 mM U-^{13}C-alanine overnight. Additionally, fresh medium containing the labelled metabolite was exchanged 2 h before metabolite extraction for steady-state analyses. To trace glucose metabolism in low-glucose conditions, cells were grown in 0.5 mM of glucose and medium was refreshed every 8 h for the 24 h labelling period to achieve steady-state labelling. For all metabolomics experiments, the quantity of the metabolite fraction analysed was adjusted to the corresponding protein concentration calculated upon processing a parallel well in a 6-cm dish.

To collect labelled conditioned medium, hPSC or 8988T cells were grown for three passages in DMEM containing 10 mM U-^{13}C-glucose, 2 mM U-^{13}C-Gln and 10% dialysed FBS. This medium was then replaced by DMEM with unlabelled glucose, glutamine and 10% dialysed FBS, and incubated for 48 h, filtered and processed for metabolite extraction. Metabolite extraction of medium was performed by adding 200 μl of filtered fresh conditioned medium to 800 μl of cold (−80 °C) methanol, incubated at −80 °C for 30 min followed by centrifugation at 10,000*g* for 10 min at 4 °C. The resultant supernatant was lyophilized by speedvac and stored at −80 °C until analysis. Dried metabolite pellets were re-suspended in 20 μl LC–MS grade water, 5 μl were injected onto a Prominence UFLC and separated using a 4.6 mm i.d. × 100 mm Amide XBridge HILIC column at 360 μl per minute starting from 85% buffer B (100% ACN) to 0% B over 16 min. Buffer A: 20 mM NH_4OH/20 mM CH_3COONH_4 (pH = 9.0) in 95:5 water/ACN. 287 selected reaction monitoring (SRM) transitions were captured using positive/negative polarity switching by targeted LC-MS/MS using a 5500 QTRAP hybrid triple quadrupole mass spectrometer.

For kinetics of metabolite secretion by hPSCs, triplicate samples of subconfluent hPSCs cultured under normal conditions were changed to fresh DMEM with 10% dialysed FBS, which was allowed to condition for 2, 4, 8, 24, 48, or 72 h. Fresh DMEM with 10% dialysed FBS was used as a blank control. Metabolites were then extracted from conditioned medium by adding ice cold 100% MeOH to a final concentration of 80% MeOH.

For PDAC metabolite uptake kinetics, conditioned DMEM with 10% dialysed FBS from subconfluent hPSCs was collected after 48 h of culture, and then filtered through a 0.45 μm filter. 8988T PDAC cells were plated in triplicate and treated with the PSC-conditioned medium or fresh DMEM with 10% dialysed FBS for 1, 2, 4, 8, or 24 h. The medium was removed and the cell lysate harvested with ice cold 80% MeOH. The soluble metabolite fractions were cleared by centrifugation, dried under nitrogen, then resuspended in 50:50 MeOH:H_2O mixture for LC–MS analysis.

For the kinetic analyses, a Shimadzu Nexera X2 UHPLC combined with a Sciex 5600 Triple TOFMS was used, which was controlled by Sciex Analyst 1.7.1 instrument acquiring software. A Supelco Ascentis Express HILIC (7.5 cm × 3 mm, 2.7 μm) column was used with mobile phase (A) consisting of 5 mM NH_4OAc and 0.1% formic acid; mobile phase (B) consisting of 98% CAN, 2% 5 mM NH_4OAc and 0.1% formic acid. Gradient program: mobile phase (A) was held at 10% for 0.5 min and then increased to 50% in 3 min; then to 99% in 4.1 min and held for 1.4 min before returning initial condition. The column was held at 40 °C and 5 μl of sample was injected into the LC–MS with a flow rate of 0.4 ml/min.

Calibrations of TOFMS were achieved through reference APCI source with average mass accuracy of less than 5 ppm except for alanine, which was 20 ppm. Key MS parameters were the collision energy and spread of 25 eV and 10 eV for positive product ion acquisition and −35 eV and 15 eV for negative acquisition. 100 MRM transitions were set on the MS method. Data Processing Software included Sciex PeakView 2.2, MasterView 1.1, LibraryView (64 bit) and MultiQuant 3.0.2.

For analysis of palmitate and stearate, PDAC cells in log growth were labelled in biological quadruplicate with either 5.5 mM U-^{13}C-glucose or 1 mM U-^{13}C-Ala in DMEM containing 2 mM glutamine and 10% dialysed FBS for 3 days. Unlabelled

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

species were used at equivalent concentrations, where relevant. Labelled medium was refreshed every day. At 72 h, medium was refreshed for 2 h, and samples were collected by quick rinse in ddH_2O followed by liquid nitrogen quenching directly on cells. Plates were then stored at −80 °C before extraction. Polar metabolites and fatty acids were extracted using methanol/water/chloroform, as described[27]. Samples were placed on ice and 10 μl of 1.2 mM D27 myristic acid as internal standard was introduced to each cell plate. 400 μl of cold water and 400 μl of methanol were added to each sample. Cells were collected in a centrifuge tube and 400 μl of ice-cold chloroform was added to each tube. Extracts were vortexed at 4 °C for 30 min and centrifuged at 14,000xG for 20 min at 4 °C. The lower (organic) phase was recovered, and samples were nitrogen dried before reconstitution in 50 μl of Methyl-8 reagent (Thermo) at 60 °C for 1 h to generate fatty acid methyl esters (FAMEs).

GC–MS analysis was performed using an Agilent 7890A GC equipped with a 30 m DB-5MS+DG capillary column and a Leap CTC PAL ALS as the sample injector. The GC was connected to an Agilent 5975C quadrupole MS operating under positive electron impact ionization at 70 eV. Tunings and data acquisition were done with ChemStation E.02.01, PAL Loader 1.1.1, Agilent Pal Control Software Rev A and Pal Object Manager updated firmware. MS tuning parameters were optimized so that PFTBA tuning ion abundance ratios of 69:219:512 were 100:114:12, increasing high ion abundance. Agilent Fiehn retention time locking (RLT) GC method was used and calibrated with standard FAMEs (Agilent) and confirmed with Agilent G1677AA Fiehn GC/MS metabolomics RTL Library. For measurement of FAMEs, the GC injection port was set at 250 °C and GC oven temperature was held at 60 °C for 1 min and increased to 320 °C at a rate of 10 °C/minute, then held for 10 min under constant flow with initial pressure of 10.91 psi. The MS source and quadrupole were held at 230 °C and 159 °C, respectively, and the detector was run in scanning mode, recording ion abundance in the range of 35–600 m/z with solvent delay time of 5.9 min. Data extraction was done with Agilent MassHunter WorkStation Software GCMS Quantitative Analysis Version B.07. Additional isotope correction was performed using an in-house software tool from MATLab[28].

All ^{13}C isotopic reagents were purchased from Cambridge Isotope Laboratories.

Quantitative PCR. Total RNA was extracted using TRIzol (Invitrogen) and reverse transcription was performed from 2 μg of total RNA using oligo-dT and MMLV HP reverse transcriptase (Epicentre), according to the manufacturer's instructions. Quantitative RT–PCR was performed with SYBR Green dye using an Mx3000PTM instrument (Stratagene). PCR reactions were performed in triplicate and the relative amount of cDNA was calculated by the comparative CT method using the 18S ribosomal or actin RNA sequences as a control.

Antibodies. LC3B (Novus Biologicals NB100-2220) was used for IF at a 1:200 dilution. Secondary anti-rabbit–GFP antibody (Invitrogen A21206) was used at 1:200. For western blot ATG5 (Novus Biologicals NB110-53818), ATG7 (Sigma A2856), β-actin (Sigma A5441), LC3B (Novus Biologicals NB600-1384), RFP (Rockland 600-401-379), and secondary HRP conjugated anti-rabbit (Thermo-Fisher, 31460) and anti-mouse (Thermo-Fisher 31430) antibodies were used, as described[14]. For IHC analysis, αSMA (Dako M0851) was used at 1:500 followed by anti-mouse–HRP secondary antibody (Vector labs PK6101).

Lentiviral mRNA targets. shRNA vectors were obtained from the RNA Interference Screening Facility of Dana-Farber Cancer Institute. The sequences and/or RNAi Consortium clone IDs for each shRNA are as follows: shGFP: GCAAGCTGACCCTGAAGTTCAT (Addgene plasmid #30323); shATG5 #1: TRCN0000150645 (sequence: GATTCATGGAATTGAGCCAAT); shATG5 #2: TRCN0000150940 (sequence: GCAGAACCATACTATTTGCTT); shATG7 #1: TRCN0000007584 (sequence: GCCTGCTGAGGAGCTCTCCAT); shATG7 #2: TRCN0000007587 (sequence: CCCAGCTATTGGAACACTGTA); shGPT1 #1: TRCN0000034979 (sequence: GCAGTTCCACTCATTCAAGAA); shGPT1 #2: TRCN0000034983 (sequence: CTCATTCAAGAAGGTGCTCAT); shGPT2 #1: TRCN0000035024 (sequence: CGGCATTTCTACGATCCTGAA); shGPT2 #2: TRCN0000035025 (sequence: CCATCAAATGGCTCCAGACAT). Mouse shATG5: TRCN0000099430 (sequence: GCCAAGTATCTGTCTATGATA); mouse shATG7 TRCN0000092163 (sequence: CCAGCTCTGAACTCAATAATA).

Chemicals. U-^{13}C-labelled glucose (Cambridge Isotope Laboratories CLM-1396-10), U-^{13}C labelled L-glutamine (Cambridge Isotope Laboratories CLM-1822-H-0.1), U-^{13}C labelled L-alanine (Cambridge Isotope Laboratories CLM-2184-H-0.1). NEAAs (Gibco 11140), D-glucose (Sigma G7528), L-glutamine (Sigma G3126), L-alanine (Sigma A7469), glycine (Sigma G8790), L-serine (Sigma S4311), sodium pyruvate (Sigma P5280), sodium L-lactate (Sigma L7022), chloroquine (Waterstone technology 32152).

Primer sequences. Sequences for qPCR primers are as follows: αSMA_Fw, GTGTTGCCCCTGAAGAGCAT, αSMA_Rv: GCTGGGACATTGAAAGTCTCA, Desmin_Fw: TCGGCTCTAAGGGCTCCTC, Desmin_Rv: CGTGGTCAGAAA CTCCTGGTT, GPT1_Fw: GTGCGGAGAGTGGAGTACG, GPT1_Rv: GATGAC CTCGGTGAAAGGCT, GPT2_Fw: CATGGACATTGTCGTGAACC, GPT2_Rv: TTACCCAGGACCGACTCCTT.

Kits. Mitochondrial stress test (Seahorse 101706-100) and glycolysis stress test (Seahorse 102194-100) kits were purchased from Seahorse Bioscience. NAD^+/NADH kit was purchased from Biovision (Biovision K337-100) and used according to the manufacturer's instructions.

Statistical analysis. Statistical analysis was done using GraphPad PRISM software. No statistical methods were used to predetermine sample size.

When comparing multiple groups with more than one changing variable (for example, experiments where cells were treated with different shRNAs and with different conditioned media) a two-way ANOVA test was performed. For experiments where we analysed one variable for multiple conditions, a one-way ANOVA was performed. In both cases, ANOVA analyses were followed by Tukey's *post hoc tests* to allow multiple group comparisons. Survival curve statistical analysis was performed using the log-rank (Mantel–Cox) test. When comparing two groups to each other, a Student's *t*-test (unpaired, 2-tailed) was performed.

Groups were considered significantly different when $P < 0.05$. The relevant calculated P values are reported in Supplementary Information, where detailed statistical information for each experiment can also be found.

Ultrasound tumour monitoring. Tumours were identified and dimensions and volume were measured as previously described using high-resolution ultrasound (Vevo 770)[29]. Briefly, mice were anaesthetized using 3% isoflurane, and abdominal fur was removed using fine clippers and depilatory cream. Pre-warmed sterile saline (100–200 μl) was administered via intraperitoneal injection. Ultrasound gel was applied over the abdominal area and the ultrasound transducer was used to identify abdominal landmark organs (liver/spleen) followed by the pancreas and the tumour. Once identified, the transducer was transferred to the 3D motor stage and a 3D scan was performed for measurement of tumour dimensions and volume. Tumour volumes were contoured as described[29].

Trypan blue-exclusion assays. To determine cell viability in starved conditions, cells were plated in complete medium at 50% confluency. Once the cells were attached, medium was replaced with serum-free DMEM. 48 h later, cells were trypsinized, re-suspended in their own medium, diluted in trypan blue (Thermo-scientific 15250061) and counted using a haematocytometer. The percentage of dead cells was determined by trypan blue incorporation.

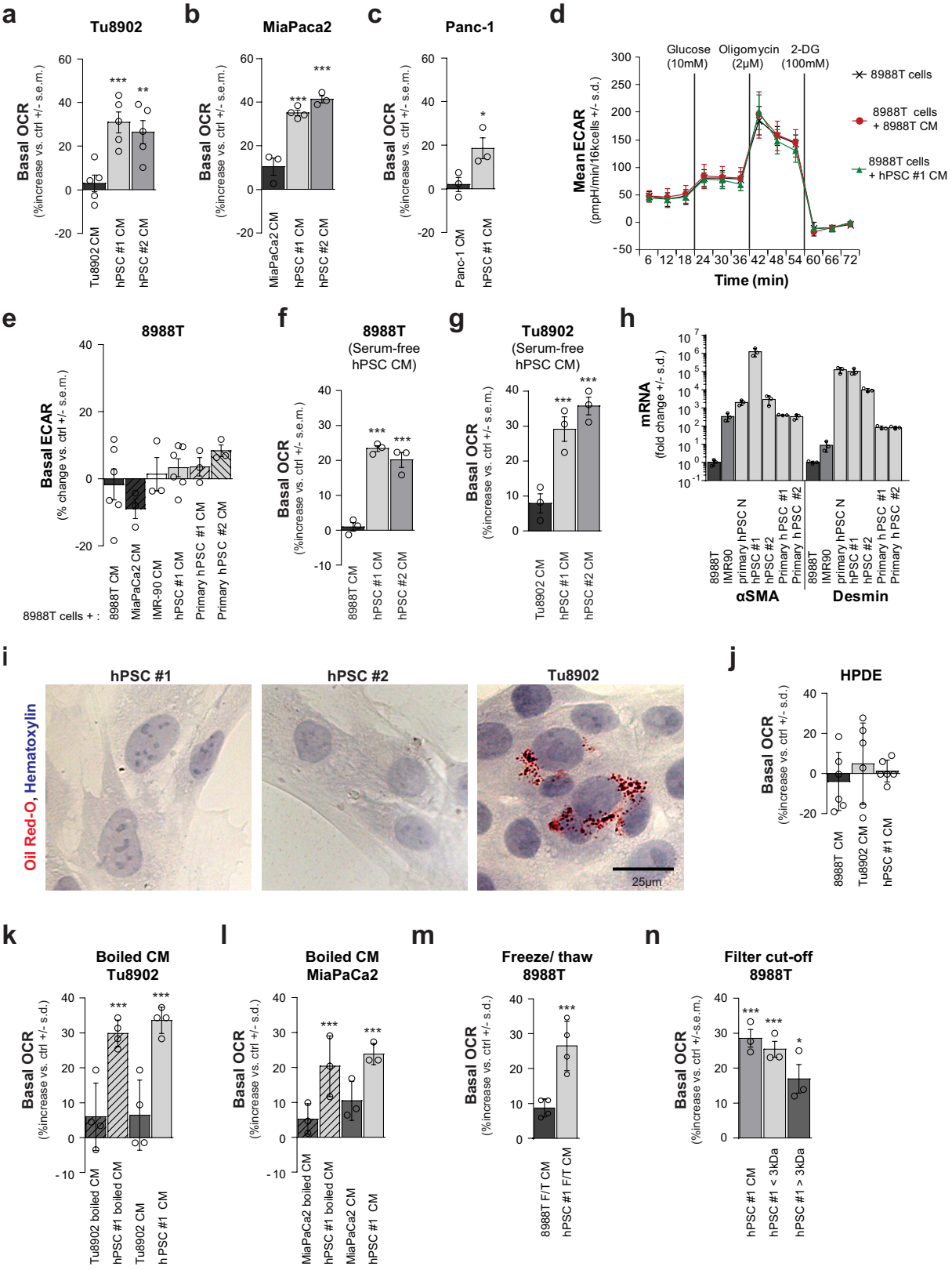
Xenografts. Xenograft studies were performed as described previously[14]. Briefly, 2×10^5 8988T or MiaPaCa2 cells were either injected alone or co-injected into the flanks of nude female mice at 6 weeks of age (Taconic ncrnu-f) with 1×10^6 hPSCs previously infected with shGFP, shATG5 or shATG7 shRNAs under protocol 10-055. Tumour take was monitored visually and by palpation bi-weekly. Tumour diameter and volume were calculated based on caliper measurements of tumour length and height using the formula tumour volume = (length × width^2)/2. Animals were considered to have a tumour when the maximal tumour diameter was over 2 mm.

For syngeneic orthotopic injections, black6 female mice at 12 weeks of age (Taconic B6NTac), pre-conditioned with doxycycline diet and kept in doxycycline regimen for the duration of the experiment, were injected in the pancreas with 1×10^5 iKRAS mPDAC cells isolated from a pure black6 PDAC GEMM (KrasG12D, P53 L/+)[15] either alone or co-injected with 5×10^5 mPSCs that were previously infected with shGFP, shATG5 or shATG7 shRNAs (or mPSC–shGFP was used alone as a negative control). Briefly, an incision was made on the flank, above the spleen. The spleen was identified and gently pulled out through the incision to expose the pancreas. 10 μl of cell suspension containing 20% of Matrigel (BD-Biosciences 354234) was injected in the tail of the pancreas using a Hamilton syringe that was held in place for 30 s to allow Matrigel polymerization. The spleen and pancreas were carefully re-introduced in the animal and the peritoneum sutured. The wound was clipped with surgical staples and the animals were allowed to recover for 1 week until the beginning of weekly ultra-sound monitoring of tumour take and progression.

Human PDAC orthotopic injections were performed in a similar way, by injecting 5×10^5 MiaPaCa2 and/or 1×10^6 hPSC #1 infected with shGFP, shATG5 or shATG7 shRNAs into the tail of the pancreas of nude female mice (Taconic ncrnu-f) at 8 weeks of age. Animals were considered as tumour-positive when a mass detected in the pancreas reached a volume of at least 1 mm^3 as calculated by 3D ultrasound. All animal studies were not blinded or randomized. Studies were performed under DFCI IACUC protocol # 10-055, where the maximal tumour size allowed is less than 2 cm.

26. Ouyang, H. *et al.* Immortal human pancreatic duct epithelial cell lines with near normal genotype and phenotype. *Am. J. Pathol.* **157,** 1623–1631 (2000).
27. Metallo, C. M. *et al.* Reductive glutamine metabolism by IDH1 mediates lipogenesis under hypoxia. *Nature* **481,** 380–384 (2012).
28. Fernandez, C. A., Des Rosiers, C., Previs, S. F., David, F. & Brunengraber, H. Correction of 13C mass isotopomer distributions for natural stable isotope abundance. *J. Mass Spectrom.* **31,** 255–262 (1996).
29. Sastra, S. A. & Olive, K. P. Quantification of murine pancreatic tumors by high-resolution ultrasound. *Methods Mol. Biol.* **980,** 249–266 (2013).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

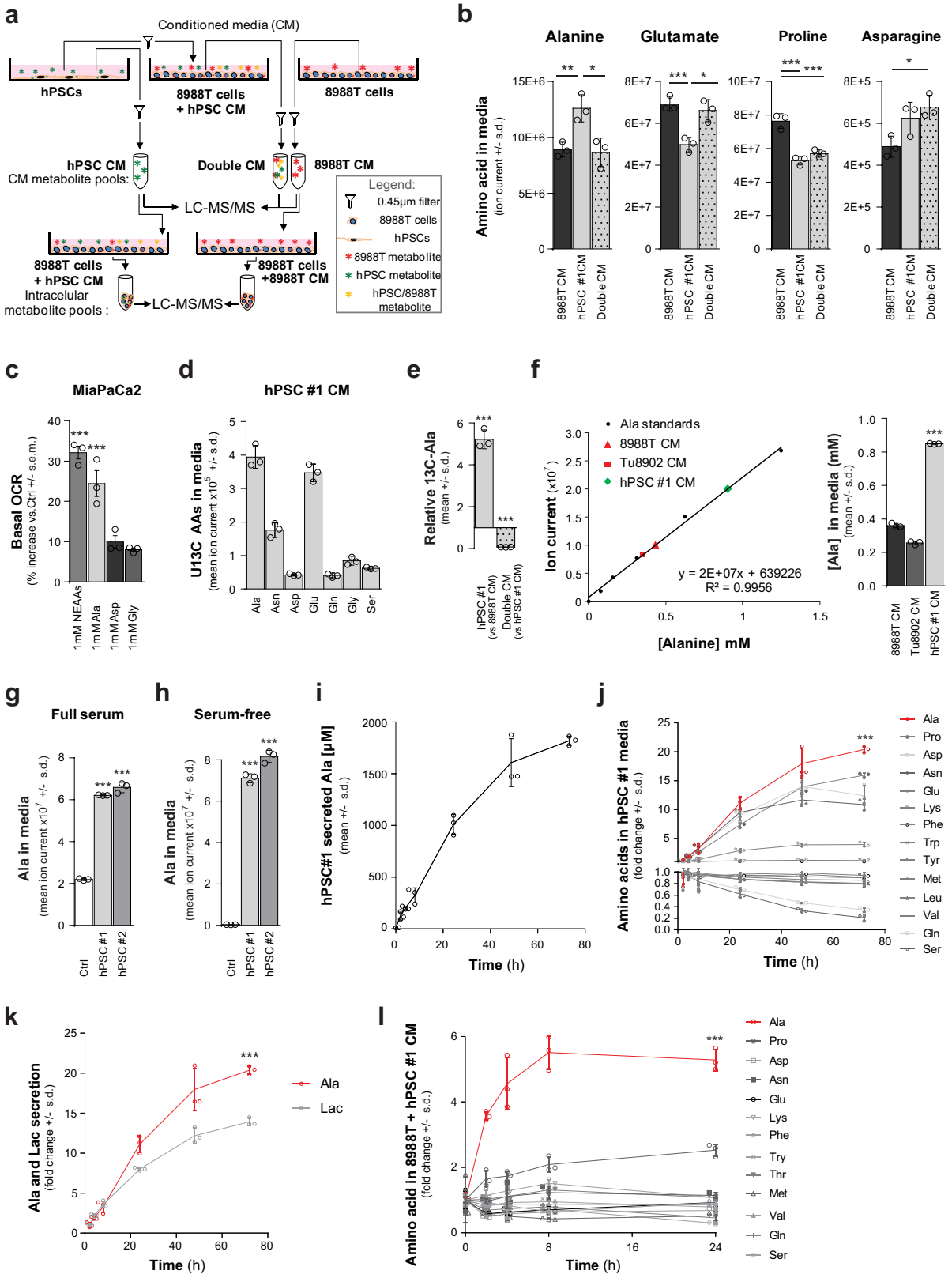


Extended Data Figure 1 | See next page for caption.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Figure 1 | Pancreatic stellate cells secrete metabolites that PDAC utilize to fuel their metabolism. a–c, Conditioned medium (CM) from human pancreatic stellate cell (PSC) lines (hPSC#1 and hPSC#2) increases oxygen consumption (OCR) in multiple PDAC cell lines: **a**, Tu8902, **b**, MiaPaCa2 and **c**, Panc-1. Data are represented as per cent increase in OCR in cells treated with conditioned medium versus cells treated with fresh DMEM containing 10% serum. Error bars represent the s.e.m. of $n = 5$ for **a** and $n = 3$ for **b**, **c**, except hPSC#1-conditioned medium in **b** where $n = 4$, from independent experiments. One-way ANOVA was performed. **a**, $***P = 0.0004$ for hPSC#1 versus control, $P = 0.0018$ for hPSC#2 versus control; **b**, $***P < 0.0001$; **c,** $*P = 0.0293$. **d**, **e**, Extracellular acidification rate (ECAR) is not significantly altered in 8988T cells when treated with conditioned medium from different cell lines. A representative Seahorse trace is shown in **d**. Error bars show s.d. of 6 independent wells from a representative tracing from 6 independent experiments (depicted in **e**). **e**, Results using conditioned medium from multiple PDAC and PSC lines including primary hPSCs. Error bars represent the s.e.m of $n = 3$ for MiaPaCa2, IMR-90, primary hPSC#1 and #2, and $n = 6$ for 8988T conditioned medium and hPSC#1, in independent experiments. **f**, **g**, Conditioned medium from hPSC#1 and hPSC#2 harvested in serum-free conditions retains the capacity to increase OCR in 8988T (**f**) and Tu8902 (**g**) cell lines. Data are represented as per cent increase in OCR in cells treated with conditioned medium versus cells treated with fresh DMEM without serum. Error bars represent the s.e.m of 3 independent experiments. 1-way ANOVA. **f**, $***P < 0.0001$; **g**, $***P = 0.0005$ for hPSC#1 versus control, $P < 0.0001$ for hPSC#2 versus control. **h**, Characterization of PSCs (primary hPSC #1 and #2 and hPSC#1 and #2) by RT–qPCR. Primary PSCs from tumours display an activated stellate cell signature in a similar fashion to that of the hPSC lines, as evidenced by the high levels of expression of the activated fibroblast marker, smooth muscle actin (αSMA), and the stellate cell marker, desmin. mRNA levels are represented as fold change compared to 8988T cells. IMR90 (human fibroblasts derived from lung tissue) and human PSCs derived from disease-free pancreata (primary hPSC N) are included as controls. Note: the transfer of normal PSCs to the tissue culture setting also leads to activation. Expression levels are normalized to β-actin. Error bars represent the s.d. of triplicate wells. **i**, Activated hPSCs are devoid of lipid droplets, an indicator of the activated state, as illustrated using oil red O staining. Tu8902 is included as a positive staining control. **j**, Conditioned medium from hPSCs does not alter OCR of non-transformed pancreatic ductal cells (HPDE). Error bars represent the s.d. of quintuplicate wells from a representative experiment (of 3 independent experiments). One-way ANOVA: $P > 0.9$. **k–m**, The ability of hPSC-conditioned medium to increase PDAC OCR is retained after boiling at 100 °C for 15 min in both Tu8902 (**k**) and MiaPaCa2 (**l**) as well as after three consecutive freeze (−80 °C, 10 min)-thaw (60 °C, 10 min) cycles in 8988T, as depicted in **m**. Error bars represent the s.d. of 4 independent wells from representative experiments (of 3 experiments). One-way ANOVA. **k**, $*P = 0.0004$ for hPSC#1 boiled conditioned medium versus control, $P = 0.0001$ for hPSC#1-conditioned medium versus control; **l**, $***P = 0.0002$ for hPSC boiled versus control, $P < 0.0001$ for hPSC versus control; **m**, $*P = 0.0011$. **n**, The factor secreted by hPSCs that increases PDAC OCR is retained in the <3-kDa fraction of hPSC-conditioned medium. Error bars represent the s.e.m of 3 independent experiments. One-way ANOVA. $*P = 0.0175$, $***P = 0.0007$ for hPSC versus control, $P = 0.0015$ for <3kDa versus control.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

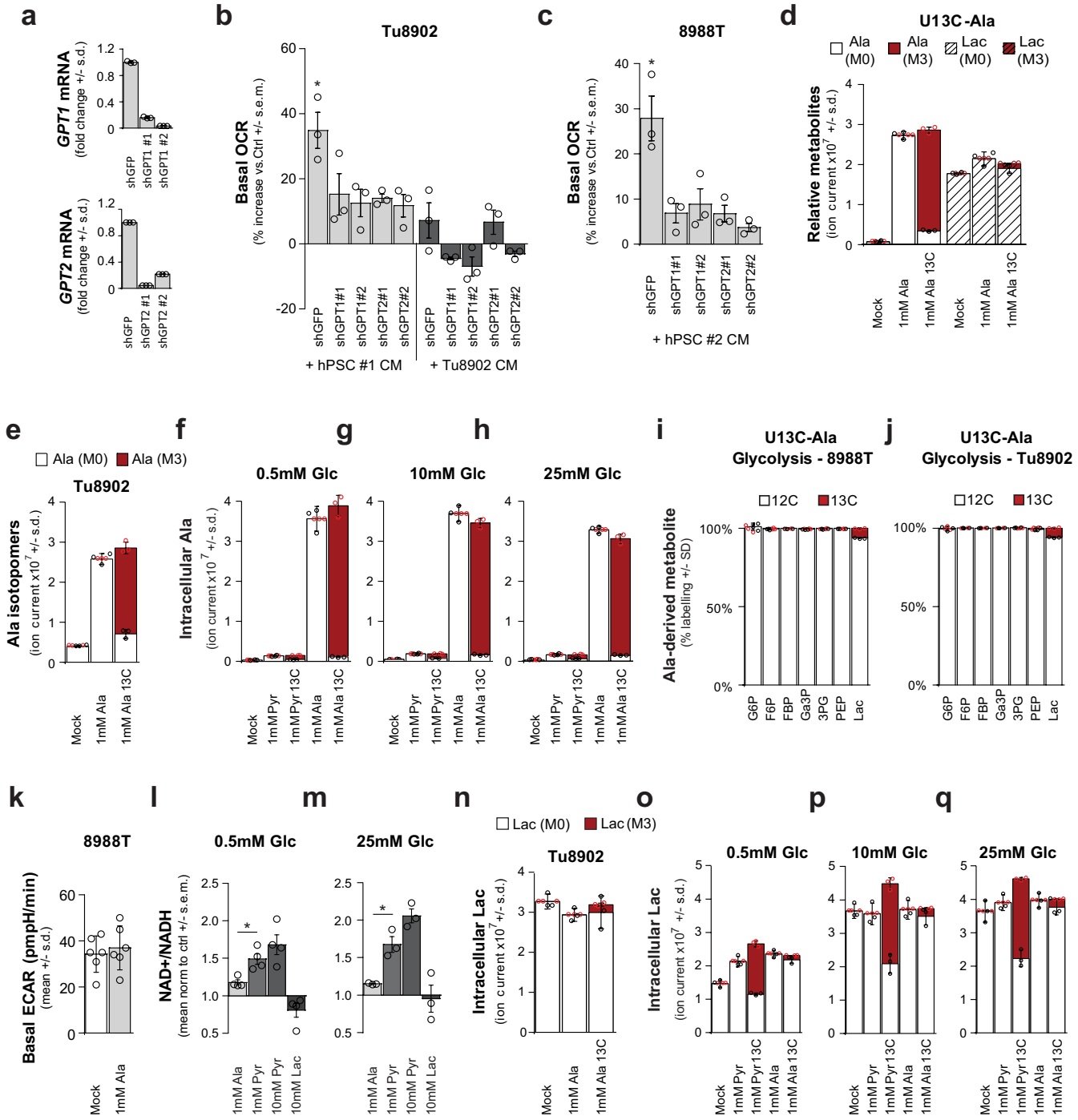


Extended Data Figure 2 | See next page for caption.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

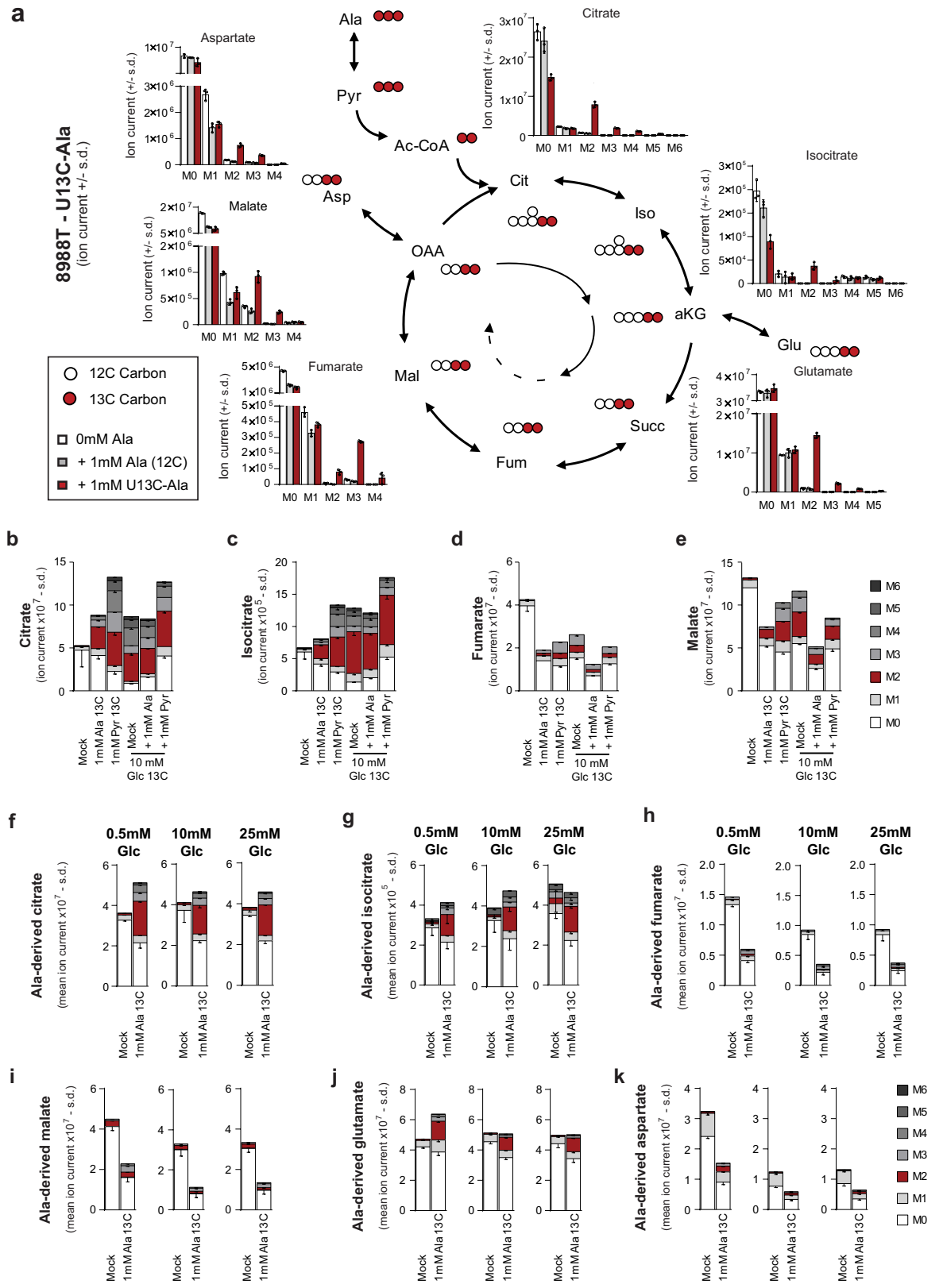
Extended Data Figure 2 | Alanine is secreted by pancreatic stellate cells and consumed by PDAC cells. a, Schematic of the metabolomic experiments depicted in Fig. 1d. **b**, The amino acids alanine, glutamate, proline and asparagine are differentially secreted by PSCs and consumed by PDAC cells. The alanine data here are also presented in Fig. 1d. Error bars represent the s.d. of $n = 3$ technical replicates from independently prepared samples from individual wells. *t*-tests were performed, for alanine: $^{*}P = 0.0176$, $^{**}P = 0.0097$; for glutamate, $^{*}P = 0.0108$, $^{***}P = 0.0024$; for proline, $^{*}P = 0.0013$ for hPSC#1 conditioned medium versus 8988T-conditioned medium, $P = 0.0024$ for double-conditioned medium versus 8988T-conditioned medium; for asparagine, $^{*}P = 0.0125$. **c**, A mixture of non-essential amino acids (1 mM of NEAA: alanine, asparagine, aspartate, glutamate, proline and serine) increases MiaPaCa2 OCR in a similar fashion to PSC-conditioned medium. Among these NEAAs, only alanine can increase PDAC OCR to an extent comparable to the NEAA mixture. Data are normalized to cells treated with fresh DMEM with 10% serum. Error bars represent the s.e.m of 3 experiments. $^{***}P < 0.0001$. **d**, Profile of the U-^{13}C-glucose and U-^{13}C-glutamine derived NEAA secretome of hPSCs. Error bars represent the s.d. of $n = 3$ technical replicates from independently prepared samples from individual wells. **e**, PSCs were labelled to saturation with U-^{13}C-glucose and U-^{13}C-glutamine. Alanine was the only labelled metabolite that showed a statistically significant increase in the PSC-conditioned medium (compared to 8988T-conditioned medium) and a decrease in the double conditioned medium (PSC-conditioned medium added to 8998T cells). Error bars, s.d. of $n = 3$ technical replicates from independently prepared samples from individual wells. A two-tailed *t*-test was performed, $^{***}P < 0.0001$. **f**, Alanine standard curve as determined by LC–MS/MS. Data points for conditioned medium from hPSC (green diamond), 8988T (red triangle) and Tu8902 (red square) are displayed on the alanine standard curve. These data are presented in Fig. 1f in μM per 10^6 cells. Error bars represent the s.d. of $n = 3$ technical replicates from independently prepared samples from individual wells. *t*-test, $^{***}P < 0.0001$. **g**, **h**, Alanine is secreted by PSCs in the presence (**g**) or absence (**h**) of serum. Error bars represent the s.d. of $n = 3$ technical replicates from independently prepared samples from individual wells. *t*-test performed; **g**, $^{***}P < 0.0001$ for control versus hPSC#1, $P = 0.0007$ for control versus hPSC#2; **h**, $^{***}P = 0.0003$ for control versus hPSC#1, $P = 0.0004$ for control versus hPSC#2. **i**, The rate of PSC Ala secretion into conditioned medium was determined over a 72-h period using LC–MS; error bars represent s.d. of $n = 4$ technical replicates from independently prepared samples from individual wells. **j**, **k**, The levels of amino acids (**j**) and lactate (**k**) in complete medium conditioned by hPSC#1 were monitored over a 72-h period. Alanine was the most relatively secreted metabolite and surpassed even lactate. Metabolite levels are normalized to time 0 (fresh DMEM with 10% dialysed serum). Error bars represent the s.d. of $n = 4$ technical replicates from independently prepared samples from individual wells. Two-way ANOVA was performed; $^{***}P < 0.0001$. The same data are used for alanine and presented in curves in **i–k**. **l**, Alanine was the most avidly consumed amino acid by 8988T cells treated with hPSC#1-conditioned medium. Error bars represent the s.d. of $n = 4$ technical replicates from independently prepared samples from individual wells. Two-way ANOVA was performed; $^{***}P < 0.0001$ for the last time-point.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 3 | Alanine secreted by stellate cells is used by PDAC to fuel biosynthetic reactions. a–c, Knockdown of GPT1 or GPT2 in PDAC cells (**a**) significantly attenuates the ability of hPSC-conditioned medium to increase OCR in Tu8902 cells (**b**). This observation was repeated with conditioned medium from an independent hPSC line in 8988T cells (**c**). Error bars represent the s.e.m of 3 independent experiments. One-way ANOVA; **b**, $*P=0.0134$; **c**, $*P=0.0129$. **d–j**, Metabolic tracing studies using U-^{13}C-Ala and U-^{13}C-pyruvate (Pyr) (**f–h**). **d**, Metabolic tracing studies using U-^{13}C-Ala in 8988T cells. Error bars, s.d. of $n=3$. Two-way ANOVA was performed: for alanine, $***P<0.0001$; for lactate, $***P=0.0001$, $**P=0.0086$. **e**, Intracellular accumulation and labelling of alanine in Tu8902 cells treated with 1 mM alanine. Two-way ANOVA was performed: $***P<0.0001$. **f–h**, Intracellular accumulation and labelling of alanine in 8988T cells treated with 1 mM pyruvate or 1 mM alanine grown in media containing different glucose (Glc) concentrations (0.5, 10, 25 mM). Error bars represent s.d. of $n=3$. **i, j**, U-^{13}C-Ala does not contribute to glycolysis or gluconeogenesis as seen for the metabolites glucose 6-phosphate (G6P), fructose 6-phosphate (F6P), fructose bis-phosphate (FBP), glyceraldehyde 3-phosphate (Ga3P), 3-phosphoglycerate (3PG), and phosphoenolpyruvate (PEP) in 8988T (**i**) or Tu8902 (**j**) PDAC cell lines. Label can be incorporated in lactate (Lac) independent of glycolysis. Error bars represent s.d. of $n=3$. **k**, 1 mM alanine does not significantly increase basal extra cellular acidification rate (ECAR) of 8988T cells. Error bars represent s.d. of 6 replicates. Data presented for a representative experiment (of 3 experiments), $P=0.6082$. **l, m**, Alanine does not alter the NAD+/NADH ratio in 8988T cells to the same extent as pyruvate in medium containing either 0.5 mM (**l**) or 25 mM (**m**) extracellular glucose. 10 mM pyruvate and 10 mM lactate were included as controls; error bars represent s.e.m of $n=4$ (**l**) or $n=3$ (**m**) independent experiments. *t*-test performed; **l**, $*P=0.0205$; **m**, $*P=0.0291$. **n–q**, Alanine contributes minimally to lactate in both Tu8902 (**n**) and 8988T (**o–q**) cells, and independently of glucose concentrations in medium, as seen by tracing of U-^{13}C-Ala: 0.5 mM (**o**), 10 mM (**p**) or 25 mM (**q**) glucose. Pyruvate labels ~50% of the lactate pool in 8988T cells (**o–q**) independent of the glucose concentration in the medium. Error bars represent the s.d. of $n=3$. **d–j, n–o**, $n=3$ technical replicates from independently prepared samples from individual wells

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

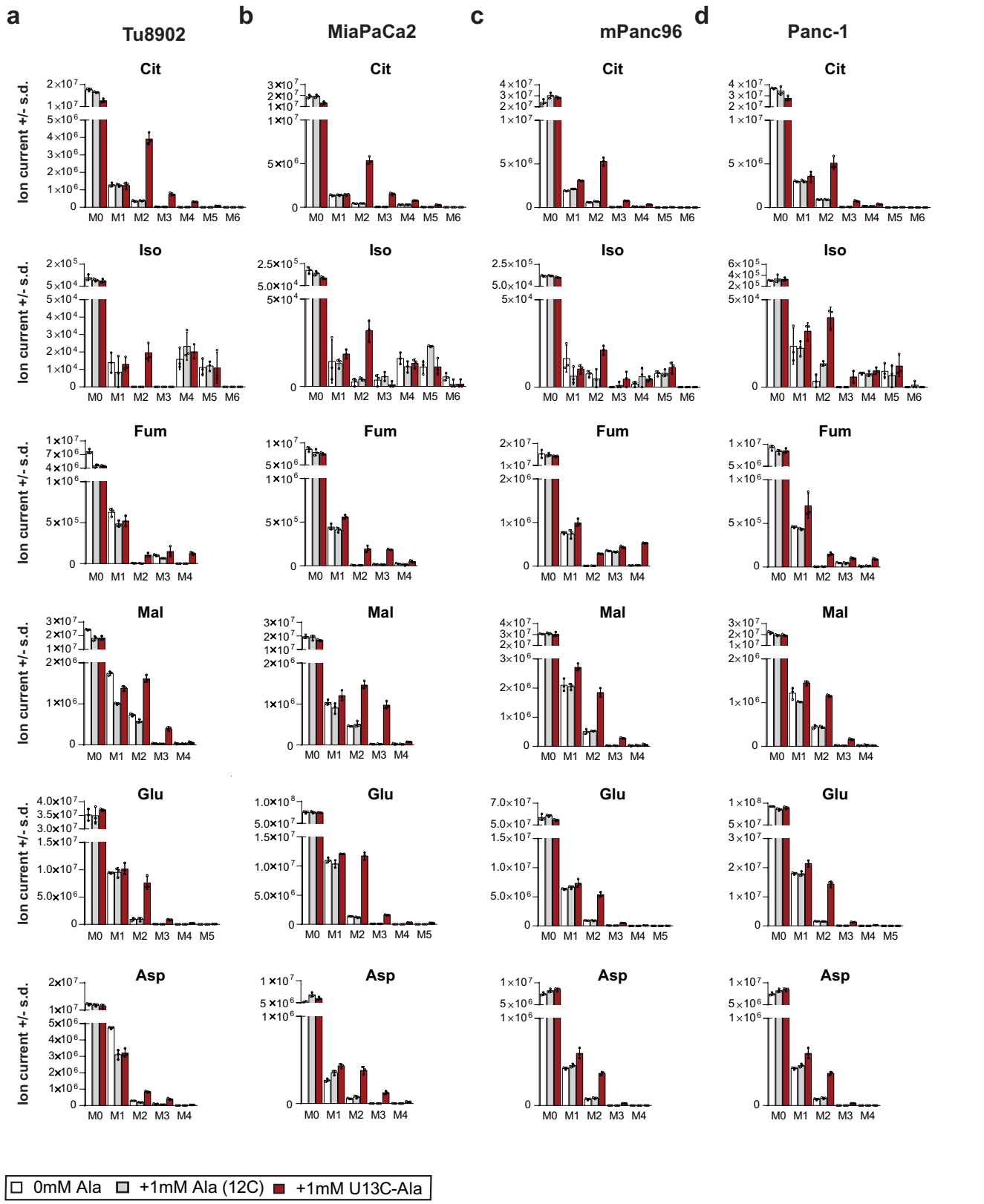


Extended Data Figure 4 | See next page for caption.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

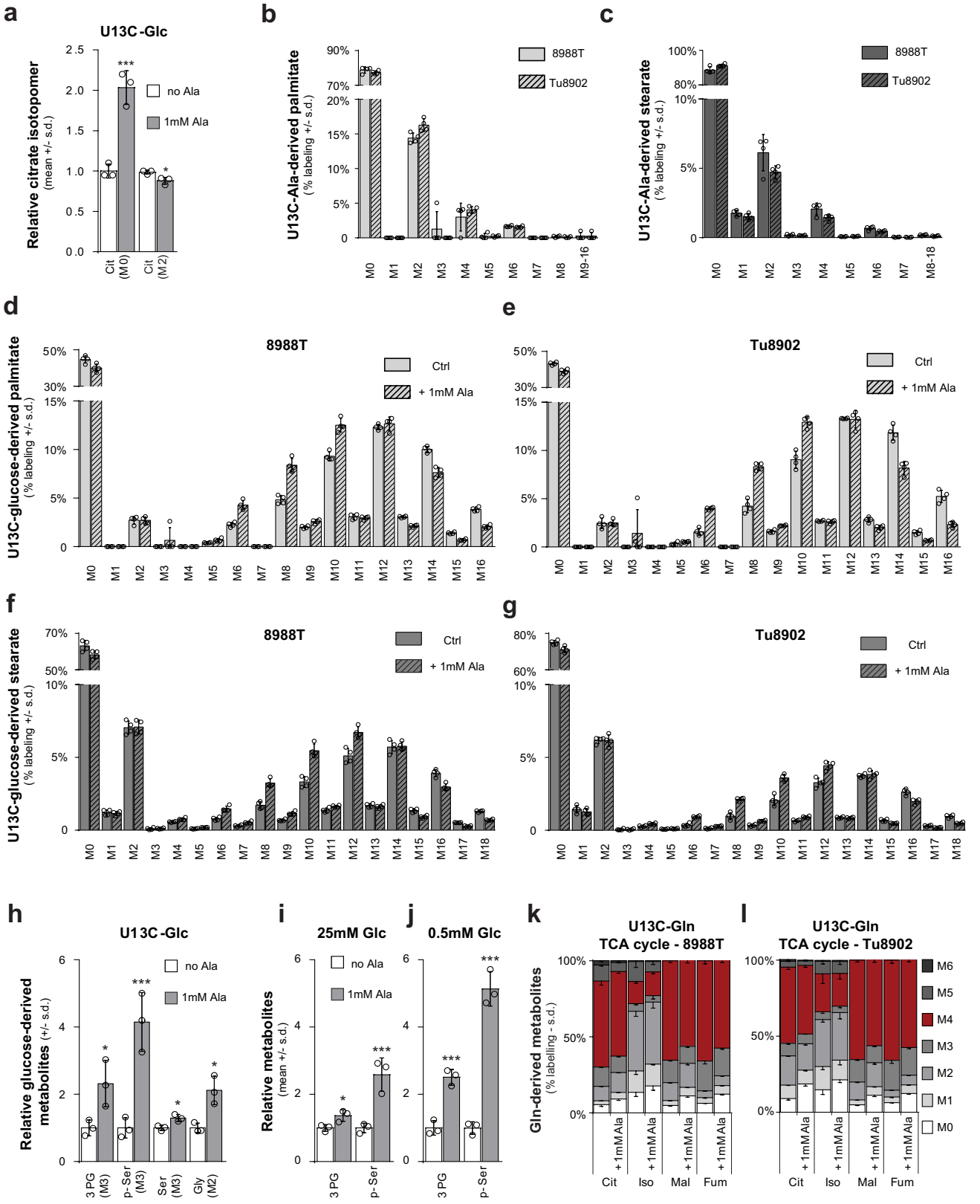
Extended Data Figure 4 | Alanine fuels the TCA cycle in 8988T PDAC cells. a, 1 mM U-^{13}C-Ala labelling of 8988T cells for 24 h shows incorporation of alanine carbon into the TCA cycle metabolites citrate (Cit), isocitrate (Iso), fumarate (Fum), malate (Mal) and NEAAs, aspartate and glutamate, derived thereof. These data are presented in Fig. 2 as fractional labelling (Fig. 2e) and as the percentage of the citrate pool incorporating the label (Fig. 2f). M0 refers to an unlabelled metabolite with no heavy carbons (unlabelled isotopomer, 12C), M1 is an isotopomer with one heavy (^{13}C) carbon that can be in any position in the molecule, M2 is a metabolite with any two heavy carbons (^{13}C), M3 with 3 and so on. The maximal M for a given species represents the fully ^{13}C labelled isotopomer (for example, for citrate that has a 6-carbon skeleton, it would be M6). In the schematic illustration, U-^{13}C-Ala is represented as three red balls, each depicting a labelled carbon atom. This is converted into U-^{13}C-Pyr (M3) and shuttled into the mitochondria. The conversion of U-^{13}C-Pyr to Ac-CoA results in the loss of one carbon as CO_2. Ac-CoA is then added to oxaloacetate (OAA) to form 2-^{13}C labelled citrate (M2). This citrate traverses around the TCA cycle and is metabolized into the other TCA cycle metabolites. The carbon labelling patterns are indicated with red (labelled) and white (unlabelled) balls. αKG, α-ketoglutarate; Succ, succinate. **b**–**e**, Contribution of alanine and pyruvate to TCA cycle metabolites citrate (**b**), isocitrate (**c**), fumarate (**d**) and malate (**e**) as shown by U-^{13}C-Ala and U-^{13}C-Pyr tracing in 8988T cells in medium containing 10 mM glucose and 2 mM glutamine. **f**–**k**, Contribution of alanine to TCA cycle metabolites citrate (**f**), isocitrate (**g**), fumarate (**h**) and malate (**i**) as well as the NEAAs Glu (**j**) and Asp (**k**) is independent of glucose concentration in the medium, as shown by U-^{13}C-Ala tracing in medium containing 0.5 mM, 10 mM or 25 mM glucose. Data are presented as total ion currents; error bars represent the s.d. of $n = 3$ technical replicates from independently prepared samples from individual wells. The raw data for **b**–**k** are presented in Supplementary Information Fig. 2a–j.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 5 | Alanine carbon contributes to the TCA cycle in multiple PDAC cell lines. a–d, 1 mM U-^{13}C-Ala labelling of Tu8902 (**a**), MiaPaCa2 (**b**), Panc-1 (**c**) and mPanc96 (**d**) cells show incorporation of alanine carbon into the TCA cycle metabolites citrate, isocitrate, malate and fumarate and the NEAAs Asp and Glu. Data are presented as total ion currents; error bars represent the s.d. of $n = 3$ technical replicates from independently prepared samples from individual wells. These data are presented in Fig. 2f as the percentage of the citrate pool incorporating label.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

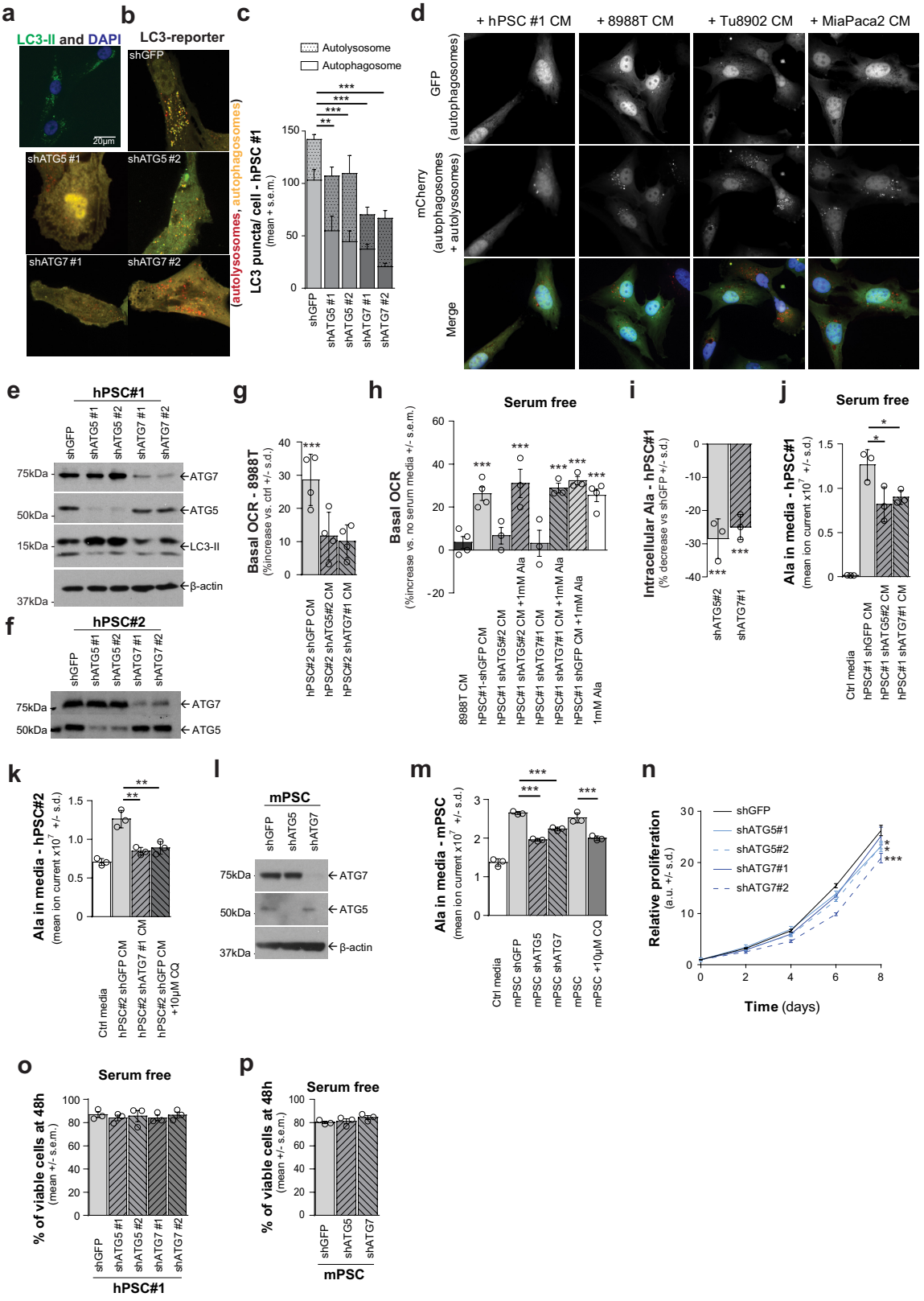


Extended Data Figure 6 | See next page for caption.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Figure 6 | Alanine relieves the demand of PDAC cells on glucose and glutamine carbon so that it can fuel other biosynthetic processes. a, The addition of alanine to PDAC cells labelled with U-^{13}C-glucose significantly increases the unlabelled citrate, with a corresponding reduction in the labelled (M2) citrate. A *t*-test was performed; ***$P = 0.0009$, *$P = 0.0169$. **b–c**, Alanine is a meaningful source of carbon for the *de novo* biosynthesis of the free fatty acids palmitate (**b**) and stearate (**c**) in two PDAC cell lines. The sum of the isotopomers from these data are presented in **Fig. 2g, h**. **d–g**, The addition of alanine to PDAC cells labelled with U-^{13}C-glucose reduces glucose carbon incorporation into palmitate in 8988T (**d**) and Tu8902 (**e**) cells as well as into stearate in both 8988T (**f**) and Tu8902 (**g**) PDAC cells. This is shown by a decrease in highly enriched species (M14, M16 for palmitate and M16, M18 for stearate) and an increase in less enriched species (M6, M8, M10 for palmitate and M6, M8, M10, M12 for stearate). **h**, U-^{13}C-glucose tracing studies illustrate that alanine drives glucose carbon into the serine biosynthetic pathway, as demonstrated by significant increases in fully labelled (M3), glucose-derived 3-phosphoglycerate (3PG), 3-phosphoserine (p-Ser), serine (Ser), and M2 glycine. A *t*-test was performed; *$P = 0.0355$ for 3PG (Ala) versus control; ***$P = 0.0041$ for p-Ser (Ala) versus control; *$P = 0.0172$ for Ser (Ala) versus control; *$P = 0.0123$ for Gly (Ala) versus control. **i**, **j**, Alanine increases serine biosynthetic pathway activity, as seen for changes in the precursors 3PG and p-Ser. This effect is enhanced in cells grown under low glucose (0.5 mM) conditions (**j**). A *t*-test was performed; i, *$P = 0.0355$, ***$P = 0.0041$; **j**, ***$P = 0.0005$ for 3PG Ala versus mock, $P < 0.0001$ for pSer Ala versus mock. **k**, **l**, Alanine alters the contribution of glutamine carbon to the TCA cycle as seen by changes in the U-^{13}C-Gln-derived fractional labelling of the TCA metabolites citrate, isocitrate, fumarate and malate upon addition of 1 mM alanine in 8988T (**k**) and Tu8902 (**l**) PDAC cell lines. Data are presented as relative metabolite for **a**, per cent labelling for **b–g, k**, **l**, and total ion currents for **h–j**. Error bars represent the s.d. of $n = 3$ (**a, h–l**) and $n = 4$ (**b–g**) technical replicates from independently prepared samples from individual wells. The raw data for **k**, **l** are presented in Supplementary Information Fig. 2k, l.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

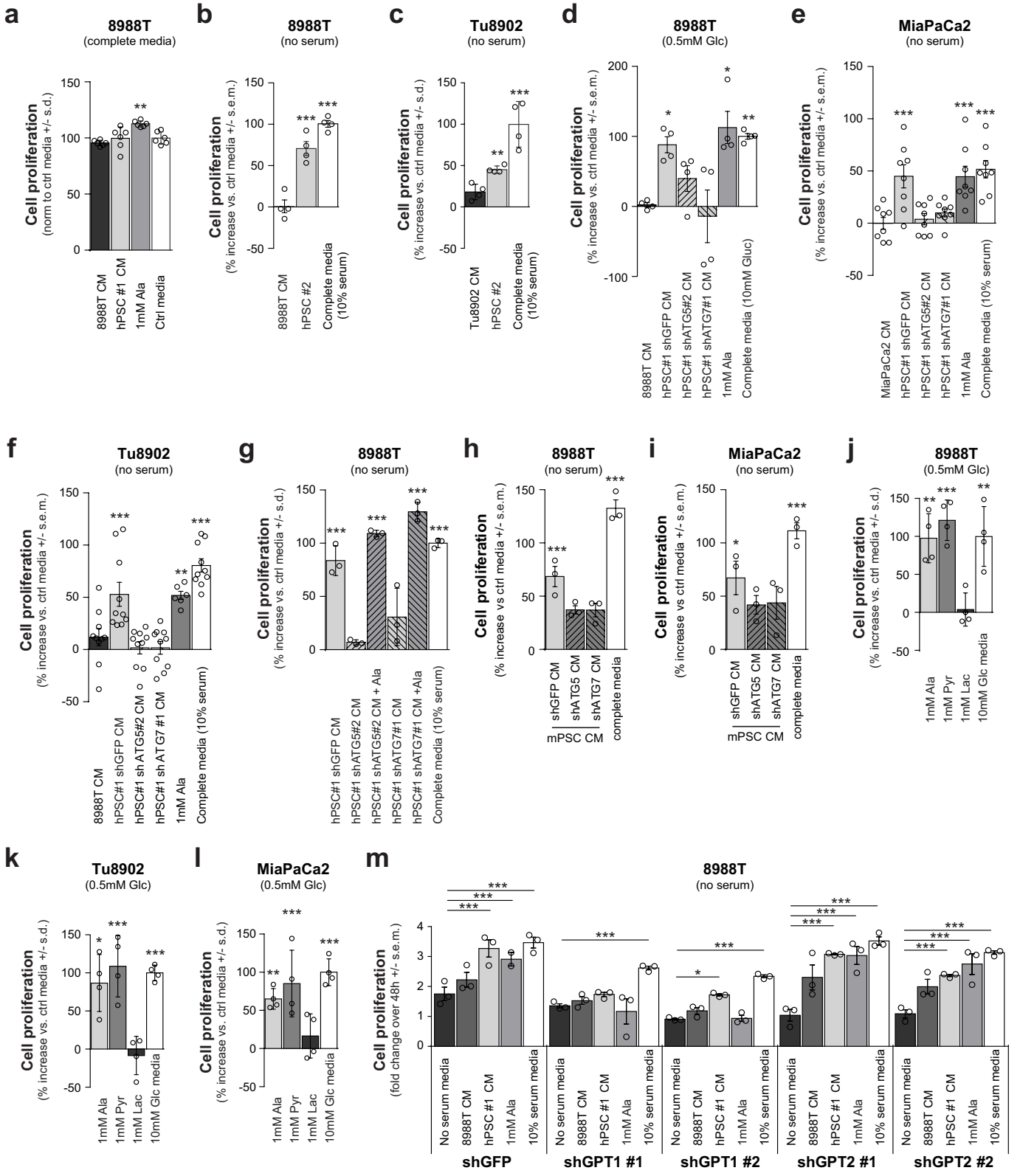


Extended Data Figure 7 | See next page for caption.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Figure 7 | Stellate cell autophagy is required to support PDAC cell metabolism through the secretion of alanine. a, PSCs immunostained for LC3-II display basal autophagy in standard culture conditions as shown by the presence of autophagosomes represented by LC3 puncta (green). Nuclei are counterstained with DAPI (blue). **b**, Representative images of autophagic puncta in control (shGFP) or autophagy impaired (shATG5 and shATG7) PSCs using an LC3 tandem fluorescence (GFP–RFP) reporter. Knockdown of *ATG5* or *ATG7* significantly decreases autophagosome formation. **c**, Quantification of autophagy in PSCs. Error bars represent the s.e.m of $n = 13$ for shGFP; $n = 12$ for shATG5#1, shATG7#1 and #2; $n = 10$ for shATG5#2. Two-way ANOVA: $^{**}P = 0.0067$ for shGFP versus shATG5#1, $^{***}P = 0.0009$ for shGFP versus shATG5#2, $^{***}P < 0.0001$ for shGFP versus shATG7#1 or #2. **d**, PDAC-conditioned medium increases autophagy in hPSC cells, as determined using an LC3 tandem fluorescence (GFP–RFP) reporter. The relative abundance of autolysosomes and autophagosomes (red and yellow puncta, respectively) is a measure of flux; data are quantified in Fig. 3c. **e**, **f**, Western blot demonstrating knockdown of *ATG5* and *ATG7* using two independent shRNAs in hPSC#1 (**e**) and hPSC#2 (**f**). A decrease in autophagy is shown by a decrease in LC3-II (lower band). **g**, Suppression of PSC autophagy by *ATG5* or *ATG7* knockdown attenuates the ability of PSC- (hPSC#2)-conditioned medium to increase PDAC OCR. Error bars represent the s.d. of quadruplicate wells from a representative experiment (of 3 experiments). One-way ANOVA; $^{***}P = 0.0006$. **h**, Suppression of PSC autophagy by *ATG5* or *ATG7* knockdown attenuates the ability of PSC- (hPSC#1)-conditioned medium to increase PDAC OCR in serum-free conditions and this phenotype can be rescued by addition of 1 mM exogenous alanine. Data are normalized to cells treated with serum-free DMEM. Error bars represent the s.e.m of $n = 4$ for 8988T, 1mM Ala, shGFP groups; $n = 3$ for shATG5#2, shATG7#1, shATG5#2 + Ala, shATG7#1 + Ala and shGFP + Ala groups in independent experiments. One-way ANOVA; $^{***}P = 0.0004$ for control versus 1 mM Ala, $P = 0.0003$ for control versus hPSC, $P < 0.0001$ for control versus hPSC shATG5 #2 + Ala, $P = 0.0002$ for control versus hPSC shATG7 #1 + Ala, $P < 0.0001$ for control versus hPSC + Ala. **i**, *ATG5* or *ATG7* knockdown in hPSCs decreases intracellular alanine concentrations compared to shGFP controls. Error bars, s.d. of $n = 3$ technical replicates from independently prepared samples from individual wells. One-way ANOVA: $^{***}P = 0.0020$ for hPSC-shGFP versus hPSC-shATG5; $P = 0.0040$ for hPSC-shGFP versus hPSC-shATG7. **j**, In serum-free conditions, *ATG5* or *ATG7* knockdown in hPSCs also decreases the secretion of alanine relative to shGFP controls. Error bars represent s.d. of $n = 3$ technical replicates from independently prepared samples from individual wells. *t*-test; $^{*}P = 0.0428$ for hPSC versus hPSC-shATG5, $P = 0.0477$ for hPSC versus hPSC-shATG7. **k–m**, Autophagy inhibition by *ATG5* or *ATG7* knockdown (**e**, **f**, **l**) or chloroquine (CQ) treatment (10 μM) decreases alanine levels in conditioned medium from hPSC#2 (**k**) and mouse PSC (**m**) as compared to shGFP or mock-treated controls. Error bars represent s.d. of $n = 3$ technical replicates from independently prepared samples from individual wells. *t*-test; **k**, $^{*}P = 0.0213$ for shGFP versus shATG7, $P = 0.02061$ for shGFP versus CQ; **m**, $^{***}P < 0.0001$ for shGFP versus shATG5, $P = 0.0003$ for shGFP versus shATG7, $P = 0.0001$ for shGFP versus CQ. **n**, Autophagy inhibition has a modest impact on PSC proliferation. Data are plotted as relative cell proliferation in arbitrary units (a.u.). Error bars, s.d. of 4 independent wells from a representative experiment (of 4 experiments). One-way ANOVA; *P* values for the last time point are as follows: $^{*}P = 0.0103$ for shGFP versus shATG5#1, $^{*}P = 0.0124$ for shGFP vs shATG5#2, $P = 0.6657$ for shGFP versus shATG7#1, $^{***}P < 0.0001$ for shGFP vs shATG7#2. **o**, **p**, Autophagy inhibition does not significantly impact hPSC (**o**) or mPSC (**p**) viability following growth for 48 h in serum-free conditions, as shown by a trypan-blue exclusion assay. Error bars represent s.e.m of $n = 3$ independent experiments.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

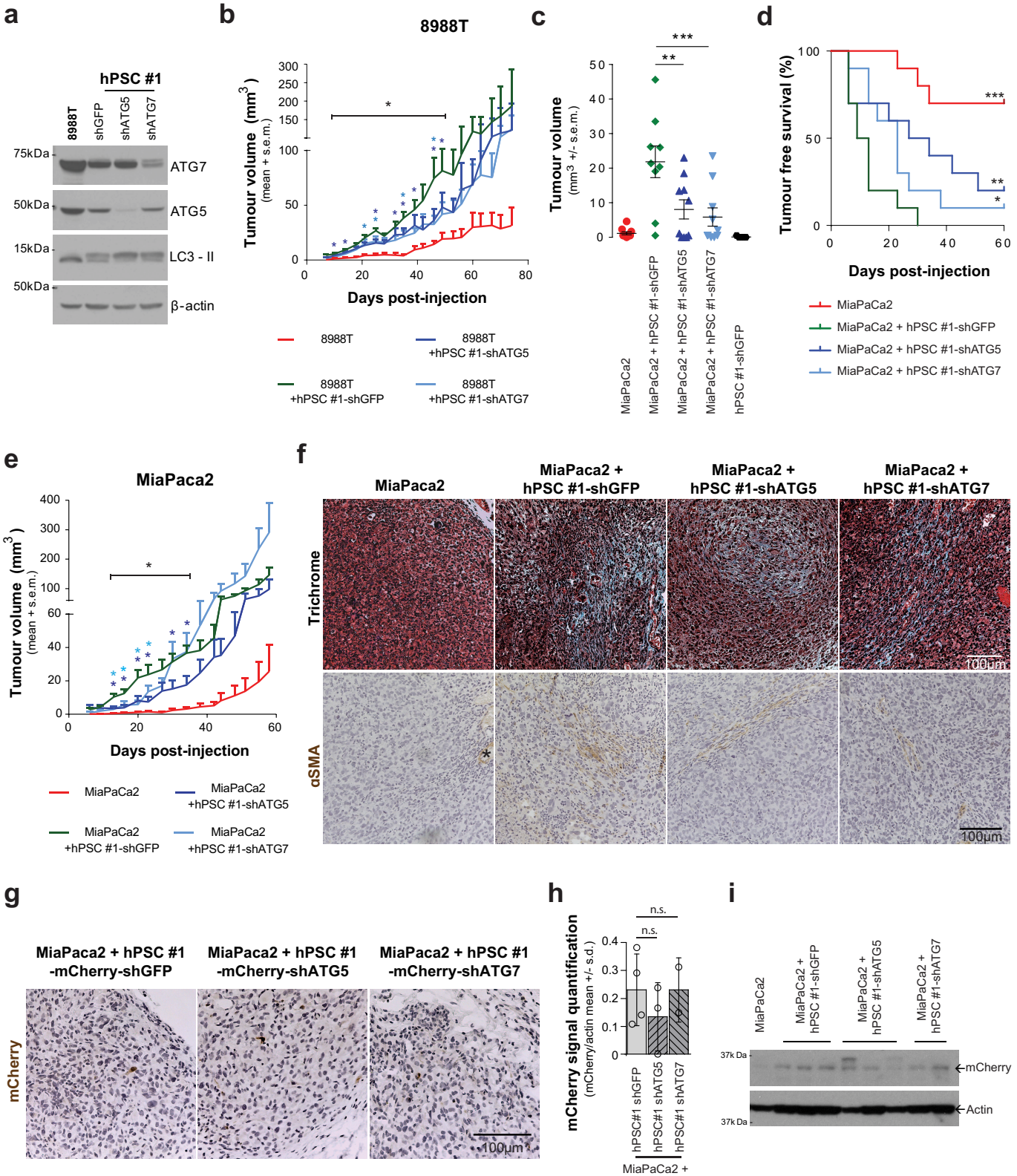


Extended Data Figure 8 | See next page for caption.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Figure 8 | Stellate cell metabolite secretion can support PDAC growth under nutrient-limiting conditions. a, hPSC-conditioned medium does not significantly affect proliferation of 8988T cells grown in complete medium (25 mM glucose, 4 mM Gln, 10% serum). Alanine supplementation contributes modestly to PDAC growth in complete medium. Error bars represent the s.d. of 6 replicate wells from a representative experiment (of 4 experiments). One-way ANOVA; $^{**}P = 0.0079$. **b**, **c**, hPSC #2 or hPSC #3-conditioned medium can sustain *in vitro* proliferation over 48 h for 8988T (**b**) and Tu8902 (**c**) cells, as compared to PDAC-conditioned medium. Data are normalized to growth in serum-free DMEM and complete serum medium is included as a positive control. Error bars represent s.e.m of $n = 4$ experiments for **b** and s.d. of quadruplicate wells from a representative experiment (of 3 experiments) for **c**. One-way ANOVA; **b**, $^{***}P < 0.0001$; **c**, $^{**}P = 0.0091$, $^{***}P < 0.0001$. **d**, hPSC-conditioned medium facilitates proliferation of 8988T cells grown in low glucose (0.5 mM) medium. Conditioned medium from hPSCs in which autophagy is inhibited loses the ability to support PDAC growth. Supplementation of medium with exogenous alanine rescues PDAC proliferation under low glucose conditions in a manner akin to the positive control, 10 mM glucose-containing medium. Error bars represent the s.e.m of 4 independent experiments. One-way ANOVA; $^{*}P = 0.0420$ for GFP versus control, $P = 0.0158$ for complete medium versus control; $^{**}P = 0.0054$. **e**–**g**, hPSC#1-conditioned medium increases PDAC proliferation in MiaPaCa2 (**e**), Tu8902 (**f**) and 8988T (**g**) cells in serum-free medium. Conditioned medium from hPSCs in which autophagy is inhibited is less effective at maintaining PDAC proliferation. Alanine increases PDAC proliferation under serum-free conditions, and it can rescue the proliferation-inducing capacity of autophagy-deficient hPSC-conditioned medium (**g**). The addition of 10% serum is included as a positive control. Error bars represent the s.e.m of 8 independent experiments for **e**, **f** and 3 independent experiments for **g**. One-way ANOVA; **e**, $^{***}P = 0.0011$ for shGFP versus control, $^{***}P = 0.0013$ for 1 mM alanine versus control, $^{***}P = 0.0001$ for complete medium versus control; **f**, $^{***}P = 0.0007$ for shGFP versus control, $^{**}P = 0.0042$ for alanine versus control, $^{***}P < 0.0001$ for complete medium versus control; **g**, $^{***}P < 0.0001$. **h**, **i**, Mouse PSC-conditioned medium increases proliferation in 8988T (**h**) and MiaPaCa2 (**i**) PDAC cell lines over 48 h in serum-free conditions. This effect is impaired when the conditioned medium is collected from mouse PSCs in which autophagy is inhibited. Error bars represent s.e.m of $n = 3$ experiments. One-way ANOVA; **h**, $^{***}P = 0.0002$ for shGFP versus control, $^{***}P < 0.0001$ for complete medium versus control; i, $^{*}P = 0.0108$, $^{***}P = 0.0002$. **j**–**l**, PDAC proliferation under low glucose (0.5 mM) conditions can be rescued by 1 mM alanine or pyruvate, but not by 1 mM lactate in 8988T (**j**), Tu8902 (**k**) and MiaPaCa2 (**l**) PDAC cell lines. Error bars represent the s.d. of quadruplicate wells from a representative experiment (of 4 experiments). One-way ANOVA; **j**, $^{***}P = 0.0013$, $^{**}P = 0.0085$ for 1 mM alanine versus control, $P = 0.0070$ for 10 mM glucose versus control; **k**, $^{*}P = 0.0323$ for 1 mM alanine versus control, $^{***}P = 0.0044$ for 1 mM pyruvate versus control, $^{***}P = 0.0010$ for 10mM glucose versus control; **l**, $^{**}P = 0.0060$ for 1 mM alanine versus control, $^{***}P = 0.0008$ for 1 mM pyruvate versus control, $^{***}P = 0.0018$ for 10 mM glucose media versus control. **m**, 8988T PDAC cells depleted for GPT1 and grown in serum-free medium do not proliferate in response to hPSC-conditioned medium or alanine, relative to control shGFP or GPT2-depleted cells. Data are represented as fold-change over 48 h and complete medium is included as a positive control. Error bars represent s.e.m. from 3 independent experiments. Two-way ANOVA; for shGFP, $^{***}P < 0.0001$ for control versus hPSC#1, $^{***}P = 0.0060$ for control versus alanine, $^{***}P < 0.0001$ for control versus complete medium; for shGPT1 #1, $^{***}P = 0.0005$; for shGPT1 #2, $^{***}P < 0.0001$, $^{*}P = 0.0493$; for shGPT2 #1, $^{***}P < 0.0001$; for shGPT2 #2, $^{***}P = 0.0004$ for control versus hPSC#1, $^{***}P < 0.0001$ for others.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

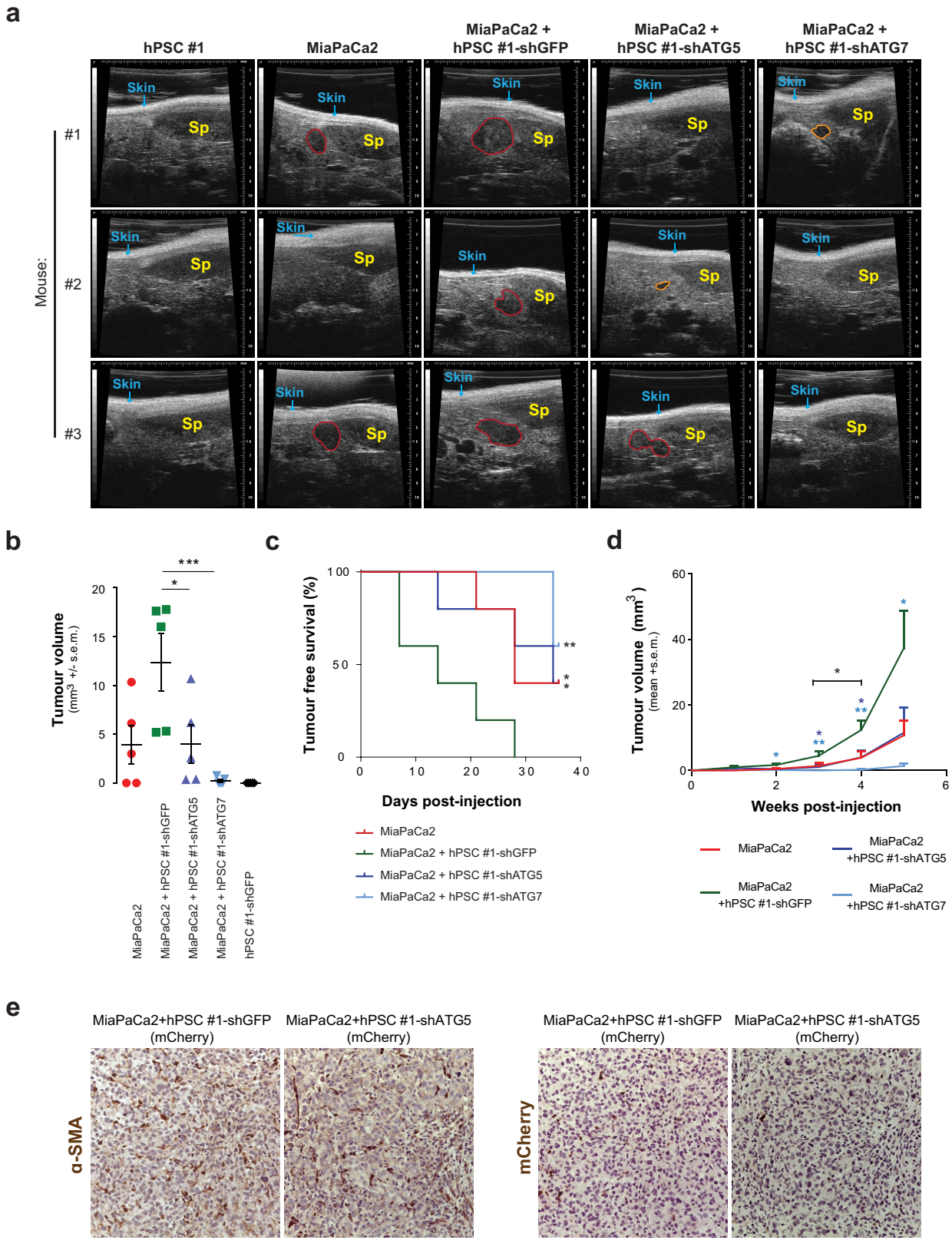


Extended Data Figure 9 | See next page for caption.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Figure 9 | Subcutaneous PDAC xenograft tumour growth is supported by autophagy-competent pancreatic stellate cells. **a**, Immunoblot for *ATG5* and *ATG7* knock-down in hPSC#1 cells infected before subcutaneous co-injection. LC3-II shows the inhibition of autophagy in these cells. **b**, Tumour growth is enhanced following co-injection of hPSCs with 8988T PDAC cells. This affect is significantly attenuated when autophagy is suppressed in the hPSCs during the initial phases of tumour growth. Error bars represent the s.e.m for 10 tumours per condition at each time point. *t*-tests were performed for each time point; $*P < 0.05$. **c–e**, Co-injection of MiaPaCa-2 PDAC cells with PSCs significantly enhances early tumour growth, analysed at 25 days post-injection (**c**) and decreases tumour-free survival (**d**). This effect is significantly attenuated when autophagy is suppressed in the PSCs. **e**, Tumour growth kinetics. Error bars represent the s.e.m. of 10 tumours per condition per time point, except for the PSC-shGFP control, for which only 5 animals were injected. **c**, One-way ANOVA, $**P = 0.0099$, $***P = 0.0020$; **d**, log-rank Mantel-Cox test, $*P = 0.0450$, $**P = 0.00174$, $***P < 0.0001$; **e**, *t*-tests for each time point, $*P < 0.05$. **f**, Representative sections for endpoint analysis from tumours for each experimental group stained with trichrome (top, blue) or α-smooth muscle actin (αSMA) (bottom, brown). Minimal residual intra-tumour collagen deposition and stromal content remains at endpoint. The asterisk in the αSMA staining images indicates a vessel and serves as a positive control. **g–i**, Early time-point analysis of MiaPaCa2 cells co-injected with RFP-labelled hPSCs in nude mouse flanks; tumours removed at 2 weeks post-injection. **g**, Representative sections of tumours stained with mCherry, a lineage label for the injected PSCs, illustrating minimal stromal content even at early time points. **h**, RFP immunoblot quantification as a marker of remaining RFP-labelled stellate cells injected. Four tumours originating from MiaPaCa2 cells co-injected with control shGFP-hPSCs, three tumours from MiaPaCa2 cells co-injected with shATG5-hPSCs and two tumours from MiaPaCa2 cells co-injected with shATG7-hPSCs were quantified. Error bars represent the s.d. of 2–4 lanes quantified per condition, as indicated. *t*-test; n.s., $P = 0.5982$ for shGFP versus shATG5, $P > 0.9999$ for shGFP versus shATG7. **i**, RFP immunoblot as a marker of remaining RFP-labelled stellate cells injected. Quantification shown in **h**. The full blot containing all nine samples is presented in Supplementary Fig. 1e.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 10 | See next page for caption.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Figure 10 | Orthotopic PDAC xenograft tumour growth is supported by autophagy-competent pancreatic stellate cells. **a**, Representative high-resolution ultrasound images of the pancreata of nude mice 4 weeks after intra-pancreatic injection with MiaPaCa2 cells alone or with control shGFP or autophagy-impaired shATG5 or shATG7 hPSCs. hPSC-only injections are included as negative controls. Skin and spleen (Sp) are indicated as spatial references; tumours are outlined in red; prospective tumours that did not meet threshold at the time of imaging are outlined in orange. **b**, **c**, Co-injection of MiaPaCa-2 PDAC cells with PSCs in the pancreata of nude mice significantly enhances early tumour growth at 21 days post-injection (**b**) and decreases tumour-free survival (**c**) in this orthotopic xenograft model. Again, this effect is significantly attenuated when autophagy is suppressed in the PSCs. Error bars, s.e.m. of 5 tumours per condition per time point. **b**, One-way ANOVA; $*P = 0.0283$, $***P = 0.0010$; **c**, log-rank Mantel-Cox test; $*P = 0.0278$ for MiaPaCa2 + hPSC#1-shGFP versus MiaPaCa2 + hPSC#1-shATG5, $*P = 0.0288$, $***P = 0.0017$ for MiaPaCa2 + hPSC#1-shGFP versus MiaPaCa2. **d**, Tumour growth kinetics following co-injection of MiaPaCa2 PDAC cells with hPSCs show enhanced tumour growth. This affect is significantly attenuated when autophagy is suppressed in the hPSCs during the initial phases of tumour growth. Error bars represent the s.e.m. of tumours from 5 animals measured per condition at each time point. *t*-tests for each time point; $*P < 0.05$. **e**, Representative endpoint sections of tumours 5 weeks post-injection stained with αSMA, a marker of PSCs, and the lineage label mCherry. Together, these stains illustrate that minimal stromal content remains at endpoint.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

LETTER

doi:10.1038/nature19107

Diverse activation pathways in class A GPCRs converge near the G-protein-coupling region

A. J. Venkatakrishnan[1,2,3,4], Xavier Deupi[5], Guillaume Lebon[6,7,8], Franziska M. Heydenreich[5,9,10], Tilman Flock[1], Tamara Miljus[5,9], Santhanam Balaji[1], Michel Bouvier[10], Dmitry B. Veprintsev[5,9], Christopher G. Tate[1], Gebhard F. X. Schertler[5,9] & M. Madan Babu[1]

Class A G-protein-coupled receptors (GPCRs) are a large family of membrane proteins that mediate a wide variety of physiological functions, including vision, neurotransmission and immune responses[1–4]. They are the targets of nearly one-third of all prescribed medicinal drugs[5] such as beta blockers and antipsychotics. GPCR activation is facilitated by extracellular ligands and leads to the recruitment of intracellular G proteins[3,6]. Structural rearrangements of residue contacts in the transmembrane domain serve as 'activation pathways' that connect the ligand-binding pocket to the G-protein-coupling region within the receptor. In order to investigate the similarities in activation pathways across class A GPCRs, we analysed 27 GPCRs from diverse subgroups for which structures of active, inactive or both states were available. Here we show that, despite the diversity in activation pathways between receptors, the pathways converge near the G-protein-coupling region. This convergence is mediated by a highly conserved structural rearrangement of residue contacts between transmembrane helices 3, 6 and 7 that releases G-protein-contacting residues. The convergence of activation pathways may explain how the activation steps initiated by diverse ligands enable GPCRs to bind a common repertoire of G proteins.

Although GPCRs have undergone extensive sequence diversification, they share a conserved structural architecture with seven transmembrane α-helices. Binding of a ligand to the receptor from the extracellular side of the cell membrane influences the rearrangement of residue contacts between the transmembrane helices[4,7]. This leads to extensive conformational changes on the cytoplasmic side of the membrane, thereby facilitating receptor activation[8–13]. The activated receptor binds to G proteins and allosterically activates them[14–17]. In this manner, the GPCR transmembrane region enables the transmission of signals across the cell membrane. At the time of writing, crystal structures have been determined for 27 class A GPCRs, belonging to seven of the eleven different subgroups (which are defined by the endogenous ligand types they bind; GPCRdb[18]; http://www.gpcrdb.org). Whereas most structures correspond to inactive states (those bound to an antagonist or inverse agonist), some structures of receptors in active states (bound to an agonist and with structural rearrangements at the G-protein-binding region) have also been solved. There are five receptors for which the structures of both inactive and active (that is, active or active-intermediate) states are available: light-activated rhodopsin[8,13], the amine-activated β_2-adrenergic receptor (β_2AR)[7], the M2 muscarinic receptor (M2R)[11], the nucleoside-activated A_{2A} receptor (A_{2A}R)[12] and the peptide-activated μ-opioid receptor (μOR)[10]. The remaining structures have been determined only in either their inactive or active states. The availability of GPCR structures from divergent subgroups (some with a sequence identity as low as around 20%[19]) that are bound to chemically diverse ligands, and known to couple to different G proteins, allowed us to investigate activation pathways across class A GPCRs.

As GPCRs are structurally similar and activate a small set of G proteins, some structural aspects of receptor activation, such as contraction of the ligand-binding site or opening of the cytosolic side owing to relocation of transmembrane helix 6 (TM6), are broadly similar. However, receptor activation is mediated by diverse ligands and therefore some aspects of ligand-associated GPCR activation must necessarily be receptor-specific. To assess the similarity of activation pathways across different receptors, we carried out a comprehensive comparison of residue contacts of inactive and active state structures. Structural equivalence for residues across the different GPCRs was assigned using the GPCR database numbering scheme[20] from GPCRdb[18] (further details in Methods). A contact between a pair of residues is said to exist if the interatomic distance between any two atoms across the residue pair is shorter than the sum of their van der Waals radii plus a cut-off distance[4,14] (Fig. 1a; further details in Methods). Analysis of residue contacts in the inactive- and active-state structures of rhodopsin, β_2AR, M2R, A_{2A}R and μOR revealed that there are 220–266 contacts between residues in the seven transmembrane helices and helix 8 for each receptor (Fig. 1b). In every receptor, roughly half of the residue contacts are reorganized upon activation.

The combinatorial possibilities of structural comparisons of multiple receptors make it difficult to determine the similarities in contacts that change upon activation among different receptors. Therefore, we used an unbiased, multi-way comparison approach in which we analysed the pattern of contacts between structurally equivalent residues across all inactive and active state structures (termed the contact fingerprint; Fig. 2a). Of the 451 contact fingerprints, only 30 represent contacts that are maintained consistently in all inactive- and/or active-state structures. The remaining 421 contact fingerprints represent contacts that are not maintained consistently in the inactive- or active-state structures across the five GPCRs. This suggests that there is marked diversity in the activation pathways of the different receptors (Fig. 2b).

Despite the diversity in the reorganization of residue contacts upon receptor activation, four contacts involving seven residues are exclusively maintained in all inactive-state structures and two contacts involving four residues are maintained exclusively in all active-state structures (Fig. 3 and Supplementary Table 1). When the roles of these contacts in the inactive- and active-state structures are considered together, a common re-organization of contacts upon activation becomes apparent, involving residue positions 3x46 in TM3, 6x37 in TM6 and 7x53 in TM7 (notation represents the GPCRdb numbering scheme) (Fig. 3). These contacts are proximal to the G protein-coupling region and therefore distant from the ligand-binding pocket. This finding highlights that the diverse structural changes among the

[1]MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, United Kingdom. [2]Department of Molecular and Cellular Physiology, Stanford University School of Medicine, 279 Campus Drive, Stanford, California 94305, USA. [3]Department of Computer Science, Stanford University, 318 Campus Drive, Stanford, California 94305, USA. [4]Institute for Computational and Mathematical Engineering, Stanford University, 475 Via Ortega, Stanford, California 94305, USA. [5]Paul Scherrer Institute, Villigen 5232, Switzerland. [6]CNRS UMR 5203, Institut de Génomique Fonctionnelle, 141 rue de la cardonille, F-34094, Montpellier, Cedex 05, France. [7]INSERM U1191, F-34094, Montpellier, France. [8]Université Montpellier, F-34094, Montpellier, France. [9]Department of Biology, ETH Zurich, Wolfgang-Pauli-Str. 27, Zurich 8093, Switzerland. [10]Department of Biochemistry & Institute for Research in Immunology and Cancer, University of Montreal, Montreal H3C 3J7, Canada.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

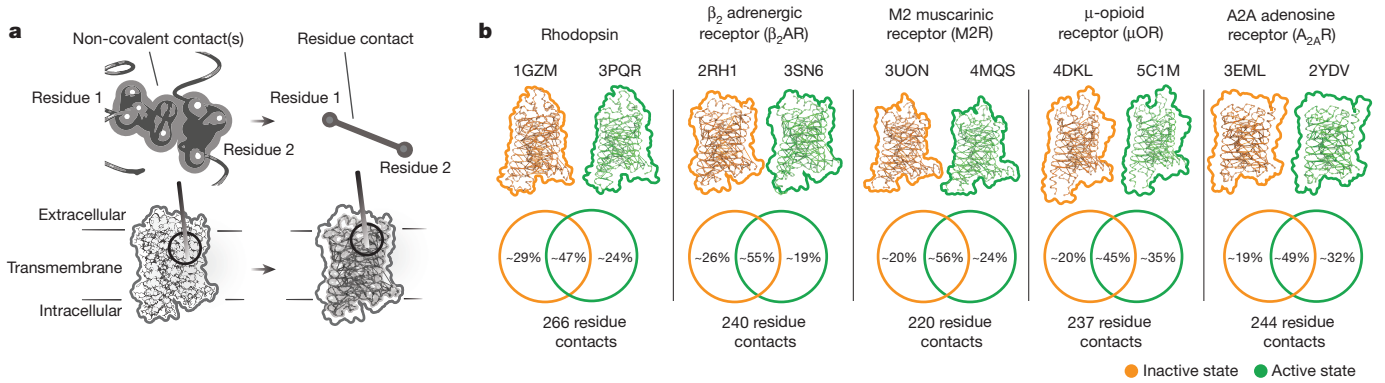


Figure 1 | Comparison of residue contacts in inactive- and active-state structures of class A GPCRs. a, Residue contact in a GPCR. Residues are denoted as circles and the non-covalent contacts between residues (residue contact) are denoted as lines connecting the circles. **b**, Similarity in residue contacts between inactive and active states of each GPCR. Comparisons of the number of residue contacts between inactive and active states are shown for five class A GPCRs using Venn diagrams.

receptors, which are stabilized by a range of different ligands, converge on a common set of rearrangements near the G-protein-coupling region.

Upon receptor activation, the cytoplasmic side of TM6 moves away from the rest of the transmembrane bundle, making several previously inaccessible G-protein-coupling residues accessible[7]. Many of these residues participate in triggering the conserved allosteric mechanism for GDP release in G proteins[14]. These include position 6x37 in the cytoplasmic end of TM6, which, upon activation, contacts a universally conserved leucine (position G.H5.25; notation taken from the Common G protein numbering scheme[14]) of the G protein in the β_2AR–G_s structure (Fig. 4a). In its inactive state, the residue at 6x37 is engaged in a residue contact with a conserved hydrophobic residue at position 3x46 (Fig. 4a). Upon activation, the residue at 3x46 breaks the contact with the one at 6x37 and forms a new contact with a tyrosine residue, Tyr7x53, within the highly conserved NPXXY motif of TM7. In the inactive state, Tyr7x53 is not available to engage with 3x46 because TM7 and TM3 are far apart and require TM6 to move out in order to form a contact. The residues at position 8x50 and 1x53 contact the residue at 7x53 in the inactive state, and both contacts are lost upon receptor activation. This contact rearrangement is observed consistently in each of the five GPCRs examined, even though they are from different GPCR subgroups, bind to different ligands and couple to different G proteins. Thus, despite the diversity in contact reorganization upon activation (Fig. 2b), there exists a common rearrangement of residue contacts near the G-protein-coupling site that underlies the activation of diverse class A GPCRs (Fig. 4a).

To assess the consistency of the contacts between residues 3x46, 6x37 and 7x53, we investigated all the other class A GPCR structures that were available only in either the inactive or activate state (further details in Methods). These include functionally diverse receptors belonging to seven of the eleven subgroups of class A GPCRs: light-activated opsins, aminergic receptors, peptide-binding receptors, protein-binding receptors, nucleoside-binding receptors, lipid-binding receptors, and others (Fig. 4b). The contact between 3x46 and 6x37 is present consistently in all inactive state structures but in none of the active state structures (Fig. 4b). Similarly, the contact between 3x46 and 7x53 is present consistently in all active-state structures but in none of the inactive-state structures (Fig. 4b). To investigate the importance of these residue positions and contacts for class A GPCRs (including in those that lack a solved structure), we performed a sequence analysis of all human non-olfactory class A GPCRs. We found that the residues that mediate contacts tend to be predominantly large hydrophobic or aromatic residues and therefore are likely to fulfil the distance criterion for contact formation (Extended Data Fig. 1). Together, these observations suggest that the rearrangement of contacts involving positions 3x46, 6x37 and 7x53 is likely to be conserved and important for receptor activation in all class A GPCRs.

In order to investigate experimentally the importance of the interactions between 3x46, 6x37 and 7x53 in a receptor for which there is no crystal structure available, we examined the signalling profile of the V2 vasopressin receptor using BRET signalling assays[21,22]. We created single-, double- and triple-alanine mutants at these positions and measured vasopressin-induced G_s and G_q activation to assess

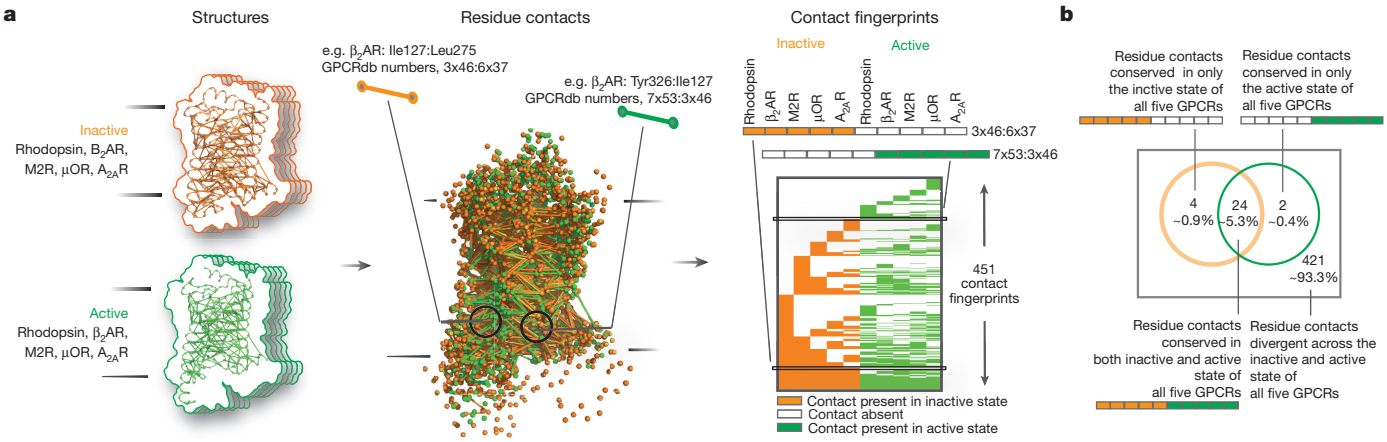


Figure 2 | Patterns of residue contacts across inactive and active state structures of GPCRs. a, For every residue contact, the presence (or absence) of a contact between structurally equivalent residues is computed for all structures. The pattern of presence and absence (filled and empty cells, respectively) across multiple structures is termed the contact fingerprint. **b**, Distribution of contact fingerprints representing the presence of contacts in the five inactive and five active state structures.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

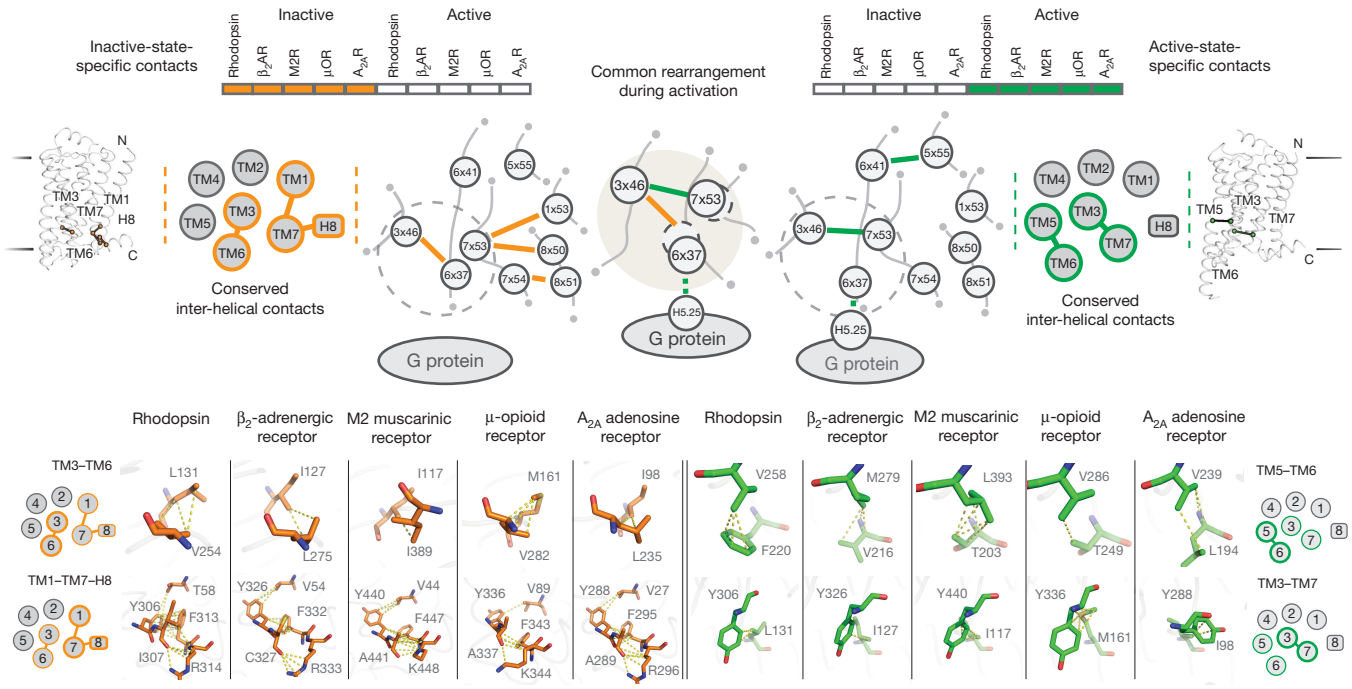


Figure 3 | Conserved rearrangement of residue contacts between inactive and active state structures of GPCRs. Considering the conserved residue contacts in the inactive (left) and the active (right) states together reveals a conserved rearrangement of residue contacts in all the five receptors. The cartoons show schematic representations of the contacts at the secondary structure level (inter-helical) and the residue level (inter-residue). Transmembrane helices and helix 8 are shown as circles and inactive- and active-state-specific contacts between the helices are shown as lines. Residues are shown as circles and contacts between residues are shown as lines. Inactive- and active-state-specific contacts are shown in orange and green, respectively. Dotted circles around 6x37 and 7x53 denote the movement of TM6 and TM7 upon activation.

receptor function. The mutations resulted in markedly reduced ability of the receptor to activate G_s and G_q when compared to the wild-type receptor (Extended Data Fig. 2). Previous observations from mutagenesis experiments in other GPCRs also highlight the importance of these positions for signalling (Supplementary Table 2). For instance, mutation of 6x37 in α_{1b}-AR results in a reduction of more than 70% in downstream signalling[23]. Similarly, mutation of Tyr7x53 in β_2AR leads to a significant reduction in G protein activation[24]. The nature of the substitution is likely to influence how the receptor is affected[25]. Given the key role of these residues, a mutation in one of these positions is likely to disrupt functional integrity, which could lead to a pathological condition in a physiological context. Indeed, mutations at these positions in different GPCRs are associated with disease (an example being hypogonadotropic hypogonadism[26]; more can be found in Supplementary Table 3). Given the importance of positions 3x46, 6x37 and 7x53 in diverse GPCRs, we anticipate that knowledge about the conserved rearrangement of contacts will guide GPCR modelling and simulations, facilitate engineering of GPCRs and help researchers to identify disease-causing mutations in other receptors.

Several studies have provided insights into distinct mechanisms of activation in individual receptors[7–13], and have described networks of contacts that are present in various combinations in different receptors upon activation[4,10,27,28]. Here we report a highly conserved rearrangement of specific residue contacts that functions as a common step in the activation pathways of diverse class A GPCRs. Because the microenvironment (that is, the surrounding residues and second shell residues) in which this rearrangement takes place diverges among receptor families, the detailed mechanism by which this common step is facilitated is likely to be distinct for different sets of GPCRs. Despite such differences, we find that the activation pathways ultimately converge onto a conserved, specific set of contact rearrangements between topologically

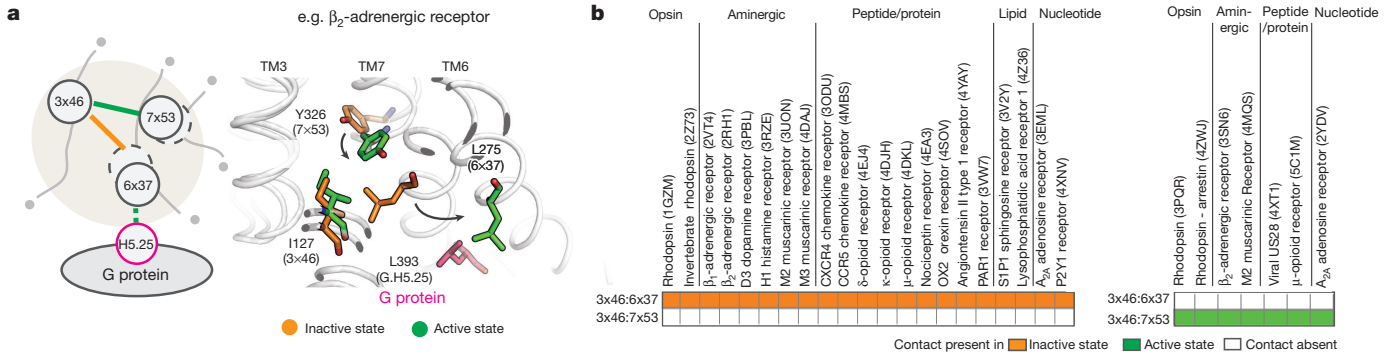


Figure 4 | Conserved rearrangement of residue contacts between inactive- and active-state structures across diverse class A GPCRs. **a**, Illustration of the conserved rearrangement of residue contacts between inactive- and active-state structures in β_2AR. Dotted circles around 6x37 and 7x53 denote the movement of TM6 and TM7 upon activation. **b**, Conservation pattern of residue contacts involved in the contact rearrangement upon receptor activation in the inactive- (left) and active-state (right) structures of class A GPCRs.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

equivalent residues near the G-protein-coupling region. Future studies aimed at investigating residues at and around these positions will help to uncover the steps that lead to the triggering of this convergent reorganization in a receptor- and ligand-specific manner.

From an evolutionary perspective, uncoupling the structural changes near the ligand-binding pocket from the G-protein-coupling region permits the ligand-binding pocket to evolve independently of the intracellular region that couples to G proteins, whilst ensuring that the activation pathways converge near the G-protein-coupling site. In this manner, the convergence of ligand-mediated allosteric activation pathways may have contributed to the evolutionary success of GPCRs: their ability to bind and be activated by a plethora of ligands, but to signal intracellularly via a small repertoire of signalling proteins inside the cell.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 30 November 2015; accepted 13 July 2016.
Published online 15 August 2016.

1. Bockaert, J. & Pin, J. P. Molecular tinkering of G protein-coupled receptors: an evolutionary success. *EMBO J.* **18,** 1723–1729 (1999).
2. Lagerström, M. C. & Schiöth, H. B. Structural diversity of G protein-coupled receptors and significance for drug discovery. *Nat. Rev. Drug Discov.* **7,** 339–357 (2008).
3. Rosenbaum, D. M., Rasmussen, S. G. & Kobilka, B. K. The structure and function of G-protein-coupled receptors. *Nature* **459,** 356–363 (2009).
4. Venkatakrishnan, A. J. *et al.* Molecular signatures of G-protein-coupled receptors. *Nature* **494,** 185–194 (2013).
5. Overington, J. P., Al-Lazikani, B. & Hopkins, A. L. How many drug targets are there? *Nat. Rev. Drug Discov.* **5,** 993–996 (2006).
6. Katritch, V., Cherezov, V. & Stevens, R. C. Structure-function of the G protein-coupled receptor superfamily. *Annu. Rev. Pharmacol. Toxicol.* **53,** 531–556 (2013).
7. Rasmussen, S. G. *et al.* Crystal structure of the β2 adrenergic receptor-Gs protein complex. *Nature* **477,** 549–555 (2011).
8. Choe, H. W. *et al.* Crystal structure of metarhodopsin II. *Nature* **471,** 651–655 (2011).
9. Dror, R. O. *et al.* Activation mechanism of the β2-adrenergic receptor. *Proc. Natl Acad. Sci. USA* **108,** 18684–18689 (2011).
10. Huang, W. *et al.* Structural insights into μ-opioid receptor activation. *Nature* **524,** 315–321 (2015).
11. Kruse, A. C. *et al.* Activation and allosteric modulation of a muscarinic acetylcholine receptor. *Nature* **504,** 101–106 (2013).
12. Lebon, G. *et al.* Agonist-bound adenosine A2A receptor structures reveal common features of GPCR activation. *Nature* **474,** 521–525 (2011).
13. Standfuss, J. *et al.* The structural basis of agonist-induced activation in constitutively active rhodopsin. *Nature* **471,** 656–660 (2011).
14. Flock, T. *et al.* Universal allosteric mechanism for Gα activation by GPCRs. *Nature* **524,** 173–179 (2015).
15. Manglik, A. *et al.* Structural insights into the dynamic process of β2-adrenergic receptor signaling. *Cell* **161,** 1101–1111 (2015).
16. Oldham, W. M. & Hamm, H. E. Heterotrimeric G protein activation by G-protein-coupled receptors. *Nat. Rev. Mol. Cell Biol.* **9,** 60–71 (2008).
17. Dror, R. O. *et al.* Signal transduction. Structural basis for nucleotide exchange in heterotrimeric G proteins. *Science* **348,** 1361–1365 (2015).
18. Isberg, V. *et al.* GPCRdb: an information system for G protein-coupled receptors. *Nucleic Acids Res.* **44** (D1), D356–D364 (2016).
19. Gonzalez, A., Cordomí, A., Caltabiano, G. & Pardo, L. Impact of helix irregularities on sequence alignment and homology modeling of G protein-coupled receptors. *ChemBioChem* **13,** 1393–1399 (2012).
20. Isberg, V. *et al.* Generic GPCR residue numbers—aligning topology maps while minding the gaps. *Trends Pharmacol. Sci.* **36,** 22–31 (2015).
21. Breton, B. *et al.* Multiplexing of multicolor bioluminescence resonance energy transfer. *Biophys J.* **99,** 4037–4046 (2010).
22. Galés, C. *et al.* Probing the activation-promoted structural rearrangements in preassembled receptor-G protein complexes. *Nat. Struct. Mol. Biol.* **13,** 778–786 (2006).
23. Greasley, P. J., Fanelli, F., Rossier, O., Abuin, L. & Cotecchia, S. Mutagenesis and modelling of the α1b-adrenergic receptor highlight the role of the helix 3/helix 6 interface in receptor activation. *Mol. Pharmacol.* **61,** 1025–1032 (2002).
24. Gabilondo, A. M., Krasel, C. & Lohse, M. J. Mutations of Tyr326 in the β2-adrenoceptor disrupt multiple receptor functions. *Eur. J. Pharmacol.* **307,** 243–250 (1996).
25. Böhm, S. K. *et al.* Identification of potential tyrosine-containing endocytic motifs in the carboxyl-tail and seventh transmembrane domain of the neurokinin 1 receptor. *J. Biol. Chem.* **272,** 2363–2372 (1997).
26. Tello, J. A. *et al.* Congenital hypogonadotropic hypogonadism due to GnRH receptor mutations in three brothers reveal sites affecting conformation and coupling. *PLoS One* **7,** e38456 (2012).
27. Isom, D. G. & Dohlman, H. G. Buried ionizable networks are an ancient hallmark of G protein-coupled receptor activation. *Proc. Natl Acad. Sci. USA* **112,** 5702–5707 (2015).
28. Lans, I., Dalton, J. A. & Giraldo, J. Helix 3 acts as a conformational hinge in class A GPCR activation: An analysis of interhelical interaction energies in crystal structures. *J. Struct. Biol.* **192,** 545–553 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank C. Chothia, G. Chalancon, N. Latysheva, R. Henderson, A. Sente, K. Kruse, A. Singhal and S. Chavali for their comments on this manuscript. We acknowledge LMB International Scholarship (in support of A.J.V.), St. John's College Benefactor Scholarship (A.J.V.), MRC Centenary Early Career Award (A.J.V.), the Medical Research Council (MC_U105185859) (M.M.B., S.B., T.F.); (MC_U105197215) (C.G.T., G.L.), Boehringer Ingelheim Fonds (T.F.), Heptares Therapeutics Ltd (G.L.), ATIP-Avenir program (G.L.), Swiss National Science Foundation (grant number 310030_153145) (G.F.X.S.); (31003A_146520) (X.D.); (31003A_159748) (D.B.V.), Swiss National Science Foundation Doc.Mobility fellowship P1EZP3_165219 (F.M.H.), the Human Frontiers Science Program (RGP0034/2014) (G.F.X.S.), the Canadian Institute for Health Research (MOP-10501) (M.B.) and the COST Action CM1207 (GLISTEN) (X.D., G.F.X.S., C.G.T., G.L.). M.M.B. is a Lister Institute Research Prize Fellow. M.B. holds a Canada Research Chair.

Author Contributions A.J.V. and M.M.B. designed the project, analysed the data, interpreted the results and wrote the manuscript, with inputs from other authors. A.J.V. collected data, wrote scripts and performed the contact fingerprinting analyses. X.D. and G.L. helped with interpretation of the activation pathways. T.F. helped with the collection of disease mutations. S.B. performed the sequence analysis of GPCRs. D.B.V., M.B., F.M.H. and T.M. designed and performed the mutagenesis and BRET assays and analysed the data. G.F.X.S. and C.G.T. provided general input on the project. M.M.B. supervised the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.J.V. (ajvenkat@stanford.edu) or M.M.B. (madanm@mrc-lmb.cam.ac.uk).

Reviewer Information *Nature* thanks D. Gloriam and the other anonymous reviewer(s) for their contribution to the peer review of this work.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Data set. The coordinates of the crystal structures of 27 GPCRs were obtained from the Protein Data Bank[29] (PDB; http://www.rcsb.org). The structures were classified into either inactive or active (active/active-intermediate) states. An inactive-state structure is one that is bound to an antagonist or an inverse agonist. An active-state structure is one bound to an agonist and has undergone structural rearrangements in the G-protein-coupling region in comparison to its inactive state structure.

There are inactive-state structures for the following GPCRs: rhodopsin (PDB accession code 1GZM), invertebrate rhodopsin (2Z73), β_1-adrenergic receptor (2VT4), β_2-adrenergic receptor (2RH1), D3 dopamine receptor (3PBL), H1 histamine receptor (3RZE), M2 muscarinic receptor (3UON), M3 muscarinic receptor (4DAJ), CXCR4 chemokine receptor (3ODU), CCR5 chemokine receptor (4MBS), κ-opioid receptor (4DJH), μ-opioid receptor (4DKL), nociceptin receptor (4EA3), δ-opioid receptor (4EJ4), orexin 2 receptor (4S0V), angiotensin II type 1 receptor (4YAY), protease activated receptor 1 (3VW7), sphingosine 1P receptor (3V2Y), lysophosphatidic acid receptor 1 (4Z36), adenosine A2A receptor (3EML), and $P2Y_1$ receptor (4XNV).

There are active-state structures for the following GPCRs: rhodopsin bound to a peptide resembling the C terminus of transducin (PDB accession code 3PQR) or bound to arrestin (4ZWJ), adenosine A2A receptor (2YDV), β_2-adrenergic receptor bound to the G protein Gs (3SN6), M2 muscarinic receptor bound to a G protein mimetic nanobody (4MQS), μ-opioid receptor bound to a G protein mimetic nanobody (5C1M), and viral US28 bound to a G protein mimetic nanobody (4XT1).

There are agonist-bound structures of four GPCRs: β_1-adrenergic receptor (PDB accession code 2Y01), serotonin 1B receptor (4IAQ), serotonin 2B receptor (4IB4) and neurotensin 1 receptor (4GRV). The β_1-adrenergic receptor structure (2Y01) is nearly identical to its inactive state structure in the G-protein-coupling region and is in an agonist-bound-inactive conformation. Since there are no antagonist- or inverse-agonist-bound structures of serotonin 1B receptor, serotonin 2B receptor and neurotensin 1 receptor, their agonist-bound structures (4IAQ, 4IB4, and 4GRV) cannot be reliably classified as in an inactive state or an active state. As a result these structures were excluded.

There are three structures of the P2Y12 receptor (PDB accession codes 4NTJ, 4PXZ and 4PY0). 4PXZ and 4PY0 are similar to each other, and have a distorted TM6 (in the cytoplasmic side) owing to the fusion with BRIL (the helix 'ends' right after 6x36, becoming a short unstructured region that links to BRIL). 4NTJ is also unusual as the fusion protein still distorts the cytoplasmic region near 6x36 but the ligand binds in a peculiar pose that forces TM6 to move away from the transmembrane bundle at the extracellular side. This modifies the rest of the helix, including the cytoplasmic side. This peculiar binding pose renders most of the extracellular domain invisible, including the conserved cysteine bridge between TM3 and extracellular loop 2. Thus, due to the specific fusion splice points (in 4NTJ, 4PXZ and 4PY0) and the peculiar binding mode of the ligand (in 4NTJ), the cytoplasmic side of TM6 in the P2Y12 receptor crystal structures appears distorted and, therefore, these structures were excluded. Similarly, TM7 of FFAR1 (4PHU) is shorter than in other class A GPCRs and appears to be truncated so this receptor was also excluded.

Calculation of residue contacts. We defined a residue contact between a pair of residues as present when the distance between any two atoms from the residue pair is less than the sum of their van der Waals radii plus a cut-off distance of 0.5 Å. Contacts involving the backbone atoms between residues that are less than four amino acids apart in the protein sequence were ignored to exclude local contacts[4,14]. We only considered contacts mediated by the residues of TM1–7 or the amphipathic helix 8 (H8) (based on the GPCRdb numbering scheme), as ligand binding primarily leads to reorganization in the transmembrane bundle, and as the loops and termini are highly variable in length, structure and intrinsic disorder among the different GPCRs[30]. While rearrangements of water-mediated contacts have been shown to be important for receptor activation[10,31], they are not considered here because not all structures have water molecules resolved in their coordinates. In each structure, residues present in the TM1–7 or in H8 are treated as nodes, and the presence of atomic contacts between pairs of residues are treated as edges connecting the nodes. We then evaluated the presence of residue contacts between topologically equivalent residues across GPCR structures. Structural equivalence was assigned using the GPCRdb numbering scheme[20] from GPCRdb[18].

Contact fingerprinting. The functional importance of a given residue contact across a group of proteins can be estimated based on the extent to which topologically equivalent contacts are maintained consistently[4,14]. For every residue contact within the ensemble, the presence or absence of an equivalent residue contact between topologically equivalent positions across the rest of the GPCR structures was recorded. This information is stored as a bit string of ones (present) and zeros (absent), which are referred to here as 'contact fingerprints'. Identifying contact fingerprints that represent consistently maintained residue contacts across and between conformational states enabled us to identify the conserved rearrangements of residue contacts in class A GPCRs. For the identification of residue contacts that are specific either to the inactive or the active states, we focused only on the GPCRs that had structures determined in both the inactive as well as active state. For the analysis of contacts across GPCRs including the inactive- and active-state-only structures (Fig. 4b), the contact distance threshold was computed based on the sum of the van der Waals radii of atoms plus multiple cut-off distance between 0.5 Å and 0.8 Å, in order to account for variation in the quality of structures.

Sequence analysis. The alignment of 311 non-olfactory, class A human GPCRs were obtained from the GPCRdb[18]. A total of 285 positions, spanning all seven transmembrane segments and H8, were aligned. The percentage presence of amino acids for residues at positions 3x46, 6x37 and 7x53 was calculated from this alignment. The set of large hydrophobic or aromatic residues was defined to include the following amino acids: Val, Leu, Ile, Met, Phe, Tyr, and Trp. The odds of finding large hydrophobic or aromatic residues at both positions that form a contact across class A GPCRs was calculated as $P/(1-P)$; where P = probability of large hydrophobic or aromatic residues in both positions that mediate a contact.

Structure analysis and visualization. Computer scripts used for identifying and comparing residue contacts in protein structures were written in Perl. For code availability, please contact the authors. Visualization of protein structures and residue interaction networks was performed using PyMOL (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.).

Signalling profile of Vasopressin V2 receptor mutants using BRET-based biosensors. *Vasopressin V2 receptor construct used in the experiments.* The human vasopressin V2 receptor construct included the N-terminal SNAP-tag for easy detection and quantification and Twin-Strep tag for optional purification in pcDNA4/TO vector (SNAP-TS-V2R). The mutations were introduced using a two-fragment PCR approach followed by Gibson assembly[32] and confirmed by sequencing.

Vasopressin V2 receptor ligands. [Arg8]-Vasopressin (Cys-Tyr-Phe-Gln-Asn-Cys-Pro-Arg-Gly-NH_2; disulfide bridge: Cys1-Cys6) was purchased from Genemed Synthesis Inc. The ligand dilutions were prepared in PBS with 0.1% (w/v) BSA. Stock solutions were stored at −20 °C and dilutions were freshly prepared for each experiment.

Biosensor constructs. For the plasmids encoding the RlucII-Gα constructs, Renilla luciferase (RlucII) was inserted into the coding sequence of human $G\alpha_s$ and $G\alpha_q$ constructs at positions 117 and 118, respectively. Gγ1 was N-terminally tagged with GFP10 as described[22]. Untagged Gβ1 was used for all experiments.

Cell culture and transfection. Human embryonic kidney (HEK) 293SL cells were transiently co-transfected with SNAP-TS-V2R, different RlucII-Gα variants, Gβ1 and GFP10–Gγ1 for G protein activation measurements. HEK 293SL cells were obtained from the laboratory of Prof. Stephane Laporte, McGill University. They are clonal isolates from HEK293 cells originally obtained from ATCC. Morphology and gene expression profiles validate their identity. Mycoplasma testing is done on a routine basis for all cell and only cells free of Mycoplasma are used for experiments. Linear 25 kDa polyethylenimine (PEI) (Polysciences Inc.) was prepared in phosphate-buffered saline (PBS) (Multicell) (PEI:DNA ratio 3:1). The cells were seeded into white Cellstar PS 96-well cell culture plates (Greiner Bio-One) at a density of 20,000 cells per well and grown for 48 h at 37 °C with 5% CO_2.

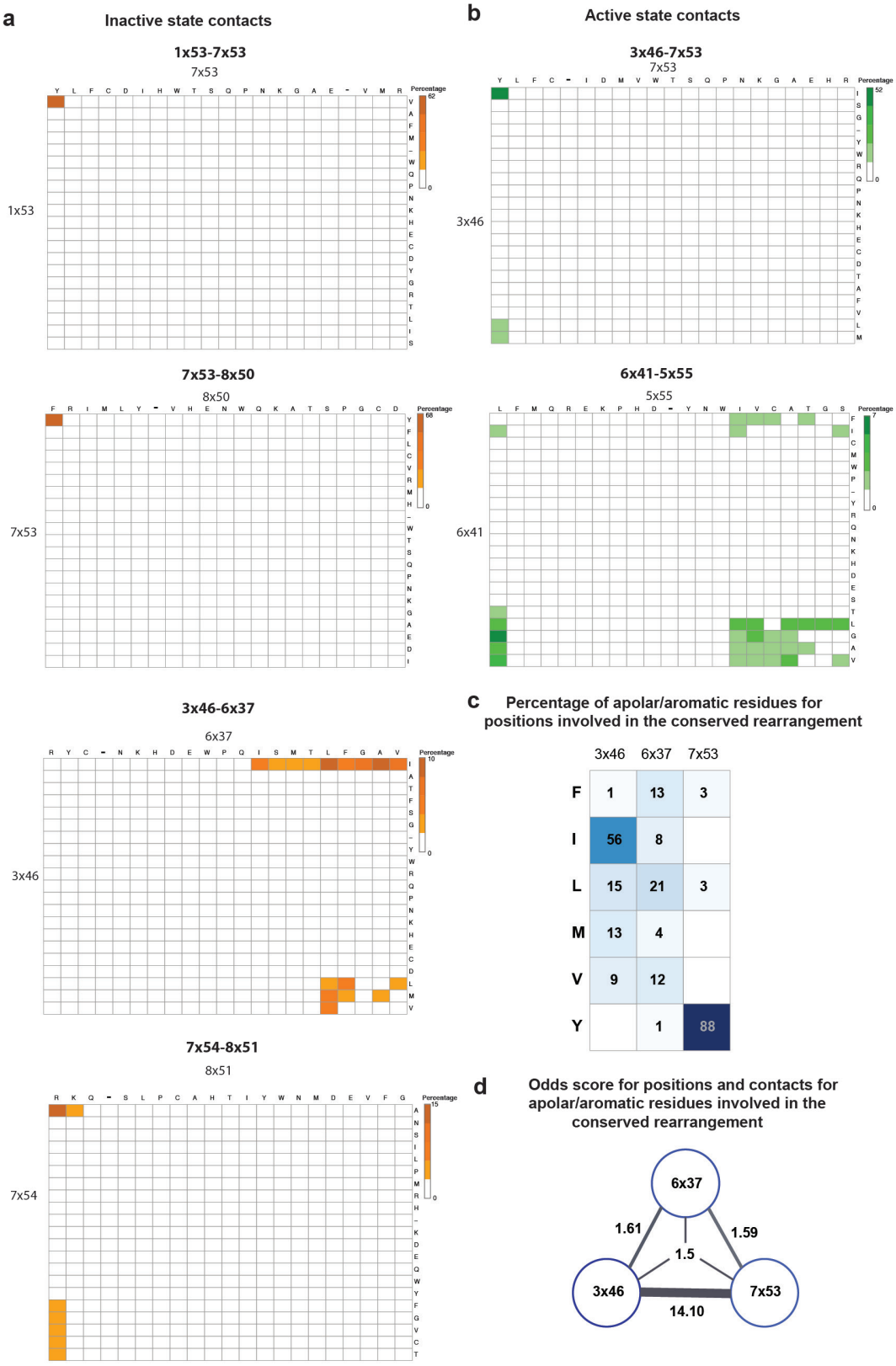
Biosensor measurements. At 48 h after transfection the 96-well plates were washed once with 200 μl PBS per well and 90 μl of Tyrode's buffer (NaCl 137 mM, KCl 0.9 mM, $MgCl_2$ 1 mM, $NaHCO_3$ 11.9 mM, NaH_2PO_4 3.6 mM, HEPES 25 mM, glucose 5.5 mM, $CaCl_2$ 1 mM, pH 7.4) were added for measurements. The cells were stored at 37 °C with 5% CO_2 for 2 h before the measurement. The plates were incubated with 10 μl ligand or vehicle per well for 5 min, then the luciferase substrate, coelenterazine 400a (DeepBlueC), was added to a final concentration of 2.5 μM. After a further 5 min of incubation, luminescence and GFP10 fluorescence were measured at 410 ± 40 nm and 515 nm ± 15 nm respectively, in a Synergy Neo plate reader (Biotek) using 0.4 s integration time. BRET was expressed as a ratio of the emission at 515 nm over the emission at 410 nm.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Quantification of expression levels. SNAP-TS-V2R wild-type- and mutant-expressing HEK293SL cells were labelled with 50 µl 1 µM SNAP-Surface Alexa Fluor 647 dye (New England Biolabs Inc.) in complete medium (DMEM, 10% FBS, Penicillin-Streptomycin) for 30 min at 37 °C, 5% CO_2 in 96-well plates. The wells were each washed with 100 µl of complete medium three times, and once with PBS. For measurements, 90 µl Tyrode's buffer were added to each well. Fluorescence was measured with 630 nm excitation and 670 nm emission.

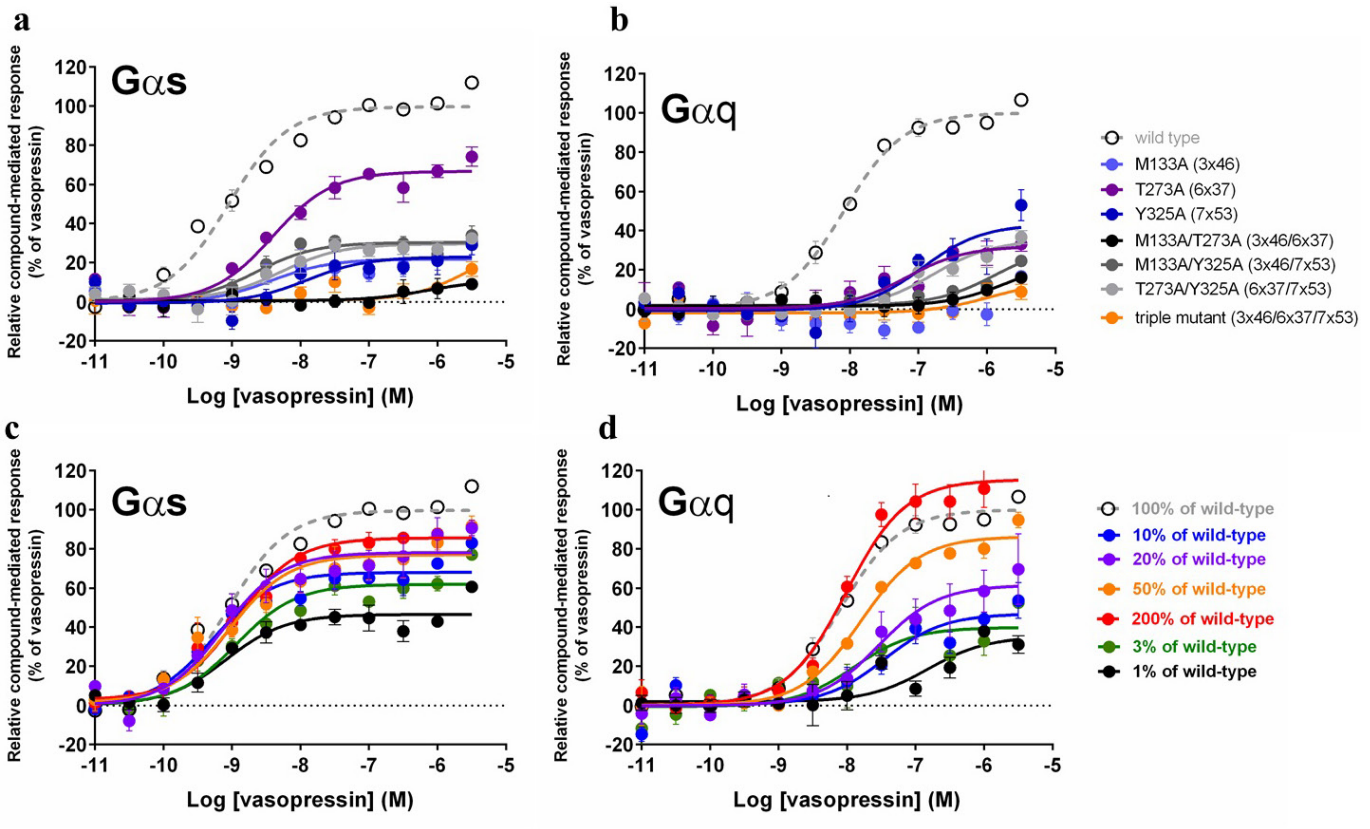
29. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28,** 235–242 (2000).
30. Venkatakrishnan, A. J. *et al.* Structured and disordered facets of the GPCR fold. *Curr. Opin. Struct. Biol.* **27,** 129–137 (2014).
31. Yuan, S., Filipek, S., Palczewski, K. & Vogel, H. Activation of G-protein-coupled receptors correlates with the formation of a continuous internal water pathway. *Nat Comms.* **5,** 4733 (2014)
32. Sun, D. *et al.* AAscan, PCRdesign and MutantChecker: a suite of programs for primer design and sequence analysis for high-throughput scanning mutagenesis. *PLoS One* **8,** e78878 (2013).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 1 | Sequence analysis of positions involved in the conserved rearrangement during receptor activation. The alignment of 311 non-olfactory class A human GPCRs were obtained from GPCRdb. Percentage of residue pairs making a contact in the inactive state only (**a**, orange spectrum; five bins) and the active state only (**b**, green spectrum; five bins). The set of large hydrophobic/aromatic residues was defined to include the following amino acids: Val, Leu, Ile, Met, Phe, Tyr, Trp. **c**, The percentage of amino acids occupying residue positions 3x46, 6x37 and 7x53 involved in the conserved rearrangement is shown. **d**, The odds of finding a large hydrophobic or aromatic residue at a pair/triplet of positions that form a contact during the conserved rearrangement. See Methods for details.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



e

construct	wild type	3x46A	6x37A	7x53A	3x46A/ 6x37A	3x46A/ 7x53A	6x37A/ 7x53A	triple mutant
logEC50 for Gs	-9.0±0.1	-8.4±0.3	-8.4±0.1	-7.9±0.3	-6.1±0.8	-8.6±0.2	-8.3±0.2	-5.5±0.9
logEC50 for Gq	-8.1±0.1	ND	-7.2±0.3	-6.9±0.3	-5.7±0.7	-5.9±0.3	-6.8±0.2	-5.9±0.6

Extended Data Figure 2 | Signalling profile of Vasopressin V2 receptor mutants using BRET-based biosensors. a, **b**, Activation of G_s (**a**) and G_q (**b**) by wild-type and mutant vasopressin V2 receptors. The residues at positions 3x46, 6x37 and 7x53 were mutated to alanine separately and in combination and their G protein signalling response was measured using BRET-based biosensors. The data points represent the mean ± SEM of $n = 3$–4 experiments. G_s is the primary and G_q the secondary cognate G protein of the vasopressin V2 receptor. For both G proteins, the activation is monitored by the decrease in BRET resulting from the separation between Gα and Gβγ. The G protein response was markedly reduced for all mutants, pointing towards reduced receptor activity. The expression levels of the mutant receptors were quantified using a fluorescent dye bound to the SNAP tag of surface-expressed receptors. While the Y325A (Y7x53A), T273A (T6x37A) and T273A/Y325A (T6x37A/Y7x53A) proteins were expressed at wild-type levels, the expression of M133A (M3x46A) and its combinations showed reduced expression levels. As a comparison, the signalling of reduced levels of wild-type vasopressin V2 receptor (% of DNA transfected) for G_s and G_q is shown in **c** and **d**, respectively. This indicates that the reduction in expression level alone cannot explain the reduced signalling of Y325A (Y7x53A), T273A (T6x37A) and the Y325A (Y7x53A)/T273A (T6x37A) combination. Thus, the reduced signalling activity of Y325A (Y7x53A), T273A (T6x37A) and T273A/Y325A (T6x37A/Y7x53A) is due to a change in the ability of the receptor to promote G protein activation. In the case of receptor mutants (single, double and triple) involving M133A (M3x46A), both the G protein response and the expression level of the receptor are markedly reduced. This may be because position 3x46 is involved in mediating contacts that are important for both the active and the inactive state, and mutating this position might simultaneously affect receptor biogenesis and downstream response. **e**, The increases in EC_{50} (along with the decrease in maximal responses) are consistent with a reduced ability of the receptor to activate G proteins.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

CORRECTIONS & AMENDMENTS

CORRIGENDUM

doi:10.1038/nature18311

Corrigendum: Cerebral cavernous malformations arise from endothelial gain of MEKK3–KLF2/4 signalling

Zinan Zhou, Alan T. Tang, Weng-Yew Wong, Sharika Bamezai, Lauren M. Goddard, Robert Shenkar, Su Zhou, Jisheng Yang, Alexander C. Wright, Matthew Foley, J. Simon C. Arthur, Kevin J. Whitehead, Issam A. Awad, Dean Y. Li, Xiangjian Zheng & Mark L. Kahn

Nature **532,** 122–126 (2016); doi:10.1038/nature17178

In this Letter we omitted to cite a relevant paper[1] showing that loss of cerebral cavernous malformation (CCM) signalling confers an increase in KLF2 expression in endothelial cells and in the developing zebrafish heart. We regret this oversight.

1. Renz, M. *et al.* Regulation of β1 integrin-Klf2-mediated angiogenesis by CCM proteins. *Dev. Cell* **32,** 181–190 (2015).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

CORRECTIONS & AMENDMENTS

ERRATUM
doi:10.1038/nature18932

Erratum: The bacterial DnaA–trio replication origin element specifies single-stranded DNA initiator binding

Tomas T. Richardson, Omar Harran & Heath Murray

Nature **534,** 412–416 (2016); doi:10.1038/nature17962

In Fig. 2c of this Letter, the rectangle at the bottom right corner was erroneously labelled 'ADP' rather than 'ATP'; this has now been corrected online.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

RETRACTION

doi:10.1038/nature18613

Retraction: Odour receptors and neurons for DEET and new insect repellents

Pinky Kain, Sean Michael Boyle, Sana Khalid Tharadra, Tom Guda, Christine Pham, Anupama Dahanukar & Anandasankar Ray

Nature **502,** 507–512 (2013); doi:10.1038/nature12594

We are retracting this Article because we no longer have confidence in data that support one of our key conclusions. In this Article we reported four advances in insect repellency: identification of olfactory neurons in *Drosophila melanogaster* that participate in repellency to *N*,*N*-diethyl-*meta*-toluamide (DEET); identification of an ionotropic receptor, Ir40a, expressed in these neurons required for avoidance to DEET; development of a chemical informatics method of identifying shared structural features from known behavioural repellents; and validation of a series of computationally identified natural chemicals as repellents for flies and mosquitoes. We no longer have confidence in data supporting that Ir40a is a DEET receptor. Upon reanalysis, the original calcium imaging (GCAMP) data show movement artefacts and background effects that we originally missed, which seriously undermine our confidence in Ir40a responses to DEET. In addition, Supplementary Fig. 5b presents several inappropriately re-used panels.

Upon learning that A. F. Silbering *et al.*[1] did not find defects in DEET aversion in Ir40a mutant flies, we repeated many of the original behaviour experiments. Although we confirmed significant behavioural differences in Ir40a cell-silenced flies (*Ir40a-Gal4;UAS-TNTG*), as reported in Fig. 2d, we have been unable to replicate observations of behavioural experiments using *Ir40a-Gal4;UAS-RNAi* flies. Therefore, with the exception of author Pinky Kain, we no longer have confidence in the conclusions of Figs 2, 3 and 5c, and Supplementary Fig. 5. We remain confident of the chemical informatics analyses and the identification of new repellents, which have been successfully repeated in our laboratory and by others, as reported in Figs 4, 5d and e, 6, and Supplementary Figs 2 and 6–9. Although it may still be possible that Ir40a does respond to DEET, given the issues listed above, all authors except Pinky Kain wish to retract this Article in its entirety. We deeply regret these circumstances and apologize to the scientific community.

1. Silbering, A. F. *et al.* Ir40a neurons are not DEET detectors. *Nature* **534,** E5–E7, http://dx.doi.org/10.1038/nature18321 (2016).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

CAREERS

PAID NETWORKING A diplomatic need for scientists **p.491**

APPLIED ABILITIES A short guide to recognizing your skills **go.nature.com/2b12gds**

NATUREJOBS For the latest career listings and advice **www.naturejobs.com**

KATE PATTERSON/*WASHINGTON POST*



Adam Kavalier developed the logo and concept for his company, Undone Chocolate, during his postdoc at Weill Cornell Medical College in New York City.

MOONLIGHTING

Dip your toes

Wayward eyes? You can explore a calling outside academia while still in the lab.

BY AMY MAXMEN

Like a marital affair, testing the waters of another career while working at the bench can feel wayward and illicit. Only later do scientists who have two-timed realize how important their digression was in shaping their career.

For some, fantasies of other careers emerge as passion for their current research fades. "I knew the evolution of this group of snails I was studying was important to some people, but I started to ask myself if it really mattered," says Stephanie Aktipis, recalling her days as a graduate student at Harvard University in Cambridge, Massachusetts. For others, academia starts to feel impractical. Forty years ago, 55% of those who gained a biology doctorate in the United States secured a tenure-track position within 6 years; by 2006, that number had dwindled to 15%, and it still seems to be dropping (see *Nature* **472,** 276–279; 2011).

Given that fall, why would a graduate student or postdoc not explore other options? "The culture in academia is to keep your head down and get your work done, but that's crazy," says geneticist Ethan Perlstein, who founded and runs a drug-discovery start-up in San Francisco, California. "You should constantly be in the process of discovery. You should think of your career like a project and experiment."

For those who do branch out, the pay-off can be life-altering. Aktipis interrupted her doctoral work to earn a master's degree in science policy, and it eventually led to her position as foreign-affairs officer at the US Department of State. Perlstein built the foundation for his own privately funded lab while doing a postdoc at Princeton University in New Jersey. He had become disenchanted with academia after receiving more than 20 rejection letters for faculty jobs.

Researchers who want to explore different options need not wait until they have finished their PhD or completed a postdoc or contract position (see 'A delicate path'). Launching a start-up, completing an internship or pursuing another degree are some of the ways in which scientists have tested the waters of a new career. To explore the entrepreneurial world, junior researchers can attend specialized ▶

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

gatherings, including those listed on meetup.com, in cities with thriving technology sectors such as Boston, London and San Francisco. Social-media platforms, particularly Twitter, can show what's happening in various start-up spheres, and provide a way to contact people with similar aspirations who could weigh in on your proposed project.

Many universities have incubator spaces where scientists can nurture start-ups. Such spaces usually offer free advice from alumni who have become chief executives, venture capitalists and lawyers. A neuroscience postdoc in the northeastern United States (who wished to remain anonymous for fear of repercussions at the university) recently used such an incubator to develop a food-related venture with three partners. Initially, they spent 10–20 hours each week on the start-up. Since incorporating the company last year, they now devote 30 hours a week to it.

The postdoc manages to juggle research and company obligations because all the experimental work is done and some flexibility is allowed in writing up publications. Once the postdoctoral stint is over, the researcher aims to focus exclusively on running the venture — a task that requires both business skills and scientific expertise. "I'm not just selling," they say. "Fifty per cent of what I do is technical and analytical."

Internships, too, provide windows into different careers, and the time commitment is more tightly controlled than it is for entrepreneurial ventures. While studying environmental microbiology as a graduate student at Harvard, Meredith Fisher began to think that she might like a science-related career in law, policy or business. She interned first with the Union of Concerned Scientists, an advocacy group in Cambridge, then with a local environmental lawyer and finally with Mass Energy, a non-profit group that promotes green energy.

Fisher created the internships by volunteering for unpaid work. Each one lasted for 2–3 months and took 8–10 hours a week. Encouraged by her experiences, she decided after earning her PhD to pursue a Master's of Business Administration, and this helped her to land her current position as a partner at a venture-capital fund in Boston, Massachusetts. "Even if you can't find internship opportunities near you," she advises, "so much can be done virtually that you can ask an organization what you might do for them remotely."

TAKING THE PLUNGE

For some, however, toe-dipping is not as effective as total immersion. That was the route taken by Aktipis. Rather than waiting to complete her PhD on snails, she took a year's leave of absence from Harvard and travelled to the London School of Economics and Political Science to do a master's degree in science policy. The programme delayed her PhD, but she learned that she loves life in the policy world.

Prasanna Bhogale took a similar approach by diving into a full-time internship. Towards the end of his PhD in physics at the University of Cologne, Germany, he was feeling unsure about a postdoc. Rather than wait to finish his PhD, he applied for internships that he found on LinkedIn and other sites.

When he got an offer for a paid six-month stint with a German investment company, he put his research on hold and accepted the post. Although he learned that finance is not his calling — he'll start work at a small analytics company in Cologne this year after completing his degree — he says that the internship transformed his thinking. "For the first time, I learned I had marketable skills," he says. "This

EASING THE WAY

A delicate path

Moonlighting requires a time commitment that is likely to be noticed. Those who have taken the plunge say that anyone considering doing likewise should consider discussing it with their principal investigator (PI). At at the very least, the PI can try to understand changes in your schedule.

For Rachel Haurwitz, co-founder and chief executive of Caribou, a start-up that is developing the genome-editing tool CRISPR, that discussion turned into a dream come true. When Haurwitz began her PhD at the University of California, Berkeley, she kept quiet about her interests in industry. But her ears pricked up whenever her adviser, Jennifer Doudna, mused about potential commercial applications of technologies developed in the lab. When those musings turned to CRISPR, the subject of Haurwitz's doctoral research, Haurwitz confessed to her interest in biotechnology. Soon the two were brainstorming names for the CRISPR company that they would co-found.

Writing CRISPR-related patent applications taught Haurwitz that she did not enjoy patent law — one career that she had quietly considered in her first year of graduate school. But she loved building the company. During her PhD programme, she took a class taught by venture capitalists, which paid off when she later became chief executive of Caribou.

Students considering jobs outside academia can learn which lab heads will be open to the idea by listening for encouraging comments about former students and colleagues now working in industry, or about discoveries being made outside the ivory tower. When Meredith Fisher was being interviewed for graduate school, she homed in on mentors who were likely to be supportive by telling every professor whom she spoke to that she did not intend to continue in academia after her PhD. Many senior faculty members indicated their disapproval, and Fisher, now a partner at a venture-capital fund in Boston, Massachusetts, ruled them out.

Some PIs expect a full commitment to lab research, and see themselves purely as mentors in academic science. Certain graduate students and postdocs say that they regret having been frank with their bosses because it cost them grant money or teaching opportunities, and it sometimes caused a personal rift. Discretion may be the best choice. A neuroscience postdoc who is nurturing a food-related start-up never works on the company in the lab, and weekly catch-ups allay any concerns that the PI might have about the progress of the research. In addition, many science departments have policies on outside employment, particularly when the university is supporting a researcher with funds. Before earning money on the side, or becoming a stakeholder in a company, students should learn about these policies from their institution's human-resources staff or office of graduate-student affairs.

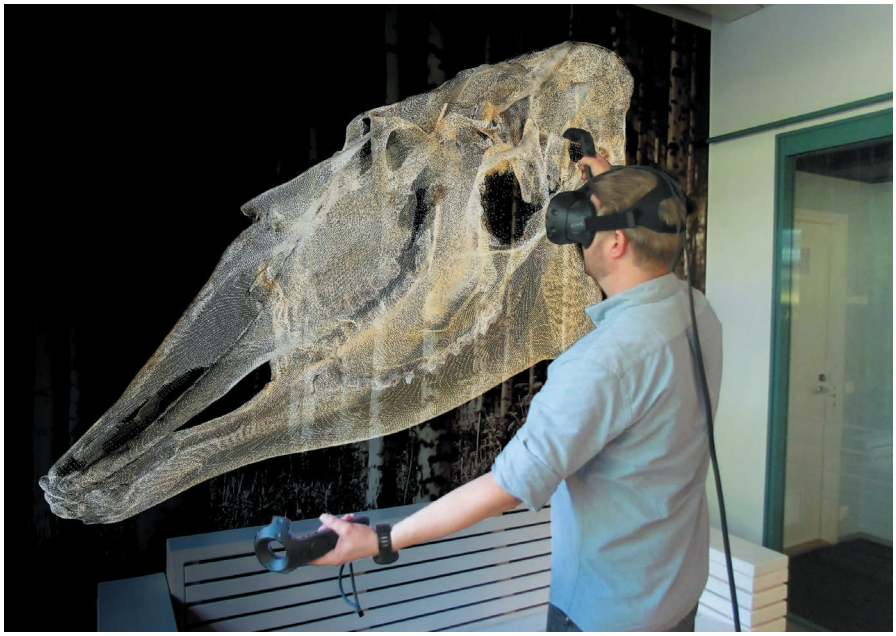
Those who decide to have 'the talk' with an adviser should be positive about their research and open about their long-term goals, Fisher says. "Rather than say, 'I don't want to do this,' you might say that your talents may lie somewhere else," she advises. Beforehand, consider how an adviser could help you, and try to recognize their needs so that you can pre-empt any concerns. And be prepared for objections. "I was asked why I wanted to waste my time with a PhD when I didn't intend to be a professor," Fisher recalls. "I would say my PhD offers the training I want in science, and the recognition it provides will be a boon to my future career." She adds: "Be proud of what you're doing." **A.M.**



Rachel Haurwitz, co-founder of Caribou.

DON FERIA/AP IMAGES FOR CARIBOU BIOSCIENCES

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



ALEKSIS KARME/TEATIME RESEARCH

Aleksis Karme is juggling his PhD in palaeontology with running his own virtual-reality company.

was psychologically hugely important, since academia can be a soul-crushing experience. It was extremely liberating to find out I was appreciated in other contexts."

SMART WORKING

Adding ten or more hours a week into an already-busy schedule is not easy, so efficiency is vital. Researchers who have successfully managed the extra workload suggest making a schedule for each day. Anything that does not contribute to the goal of the research should be cut. Social media, e-mails, phone calls and other potential time-wasters should be restricted to certain hours.

Even so, scientists who are committed to a side pursuit say that they sacrifice their social life, and sometimes their rest. "I didn't do much else besides my research and making chocolate," says Adam Kavalier, a chemist who developed the logo and concept for his company, Undone Chocolate, during his postdoc at Weill Cornell Medical College in New York City. "I didn't sleep much: I worked weekends and nights. I did not take vacations," he says. "But I already had a passion for making chocolate, and once I decided to make it a business, that became an obsession."

"For the first time I learned I had marketable skills. This was hugely important."

Passion also drew Aleksis Karme, a PhD student in palaeontology at the University of Helsinki, to his test project. He co-founded a virtual-reality company, and is now juggling that with his dissertation.

He admits to overworking, but the upside is that he can apply techniques that he creates for his lab work to real-world projects in construction and engineering. "I can control my own career better this way," he says. "I'm finding a way around the conveyor belt from a PhD to a postdoc."

NO REGRETS

Many researchers who eventually left academia for other jobs they'd sampled while at the bench say that they published less than they might have done had they focused on science alone. But that never worried them, because they knew that it would not matter in the long run. Meanwhile, those who stayed in academia say that they do not regret temporarily veering from the path.

As a postdoc, microbiologist Robin Kodner served on the board of a biofuel start-up and consulted for the biofuel industry. The experience assured her that academia was the right choice for her, and she is now a faculty member at Western Washington University in Bellingham. "Venture capitalists would ask me if I could have a product in two years, and I was like, are you joking? Being from academia, I was totally comfortable saying, 'Well, we don't really know that yet, and here are the caveats on what we do know'," she says. "But that's not how you talk with investors. I wasn't a good fit. I'm still so glad I did it because I got a taste of what that life was like, and at the same time I didn't really lose traction on my academic career."

Junior researchers need to remember that looking elsewhere is not cheating. "My advice is that you have control over your life," Aktipis says. "Know you can do this. Know it is an option." ■

Amy Maxmen *is a freelance writer in Berkeley, California.*

TRADE TALK
Binational liaison



After completing a PhD and postdoc in neuroscience, Sabine Blankenship took a job as the science liaison officer at the German Consulate General in San Francisco. She describes how she helps her compatriots to stay up to date with scientific developments across northern California and the Pacific Northwest.

What do you do?

I build professional networks, engaging researchers and learning about scientific developments. Recently, I participated in an industry day at Lawrence Berkeley National Laboratory, California. Before that, I spoke to relevant researchers to learn about the CRISPR–Cas gene-editing technology and wrote a report for German government officials. When German delegations visit, I organize their agendas. I can spend a lot of time at the computer requesting meetings, but our work depends on getting out and getting to know people. The communicative part is something I totally enjoy — that and the ability to keep learning.

How do you apply your training?

Very rarely is it factual knowledge. Anything in life sciences, I can follow in depth, but I'm also responsible for energy and information technology. What I really learned in graduate school was how to research something deeply and structure that information into a useful format. Also, self-management is important. I have to take the initiative and see that my projects keep running.

How did you get the job?

What tipped me to this position was a conference run by the German Academic International Network, which helps German nationals to find jobs back home. I thought it was the wrong conference to go to since I wanted to stay in the United States, but then I found the job ad for the consulate. Friends introduced me to two people who work at consulates and I had coffee and lunch with them. Talking to them made me realize that consulate work is something I could be good at and love. And it proved to be useful for preparing for the job interview. ■

INTERVIEW BY MONYA BAKER

This interview has been edited for length and clarity. See go.nature.com/2auauki for more.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

INTERDIMENSIONAL TRADE BENEFITS

A powerful argument for cooperation.

BY BRIAN TRENT

"It was a missile," she explained.

The admirals of the Brightworld Imperium blinked from where they sat around the crisis room. "What do you mean?" one of them said at last. "A missile? From where? From whom?"

Dr Harshadi Hennig stood at the centre of the rounded chamber, feeling less like a scientist giving a presentation and more like a condemned prisoner. She raised her holo-gloves, conjuring an image of Eris Station orbiting Neptune's serene planetscape.

"This is a recording from an approaching shuttle," Hennig said.

The Neptunian station was a grey wheel against the void. The time index ticked away until — at the 18:43:30 mark — a greenish light split the blackness. It seemed as though a massive Venetian blind was opening, an emerald sky materializing against cold space. A colossal, spire-like shape appeared. Then the light blinked out, the odd shape disappearing with it. The shuttlecam trembled in aftershock. The orbital station was —

"Gone," one of the admirals snapped in the impatient tone used for all plebeians of the Empire. "Yes, Dr Hennig, *we know this*. Tell us what the devil happened to Eris Station!"

She bowed obediently. "Of course, Admiral. The station was destroyed by a missile fired from another dimension."

"It … what?"

"A missile. Not fired by insurrectionists or terrorists, but from a universe next door."

She conjured a still-frame of the greenish rupture with her gloves.

"It appears," she continued, "that the multiverse theory is correct. This video shows the splitting of a dimensional membrane, and the brief materialization of what we believe is the alien equivalent of an ICBM."

The admiralty board conferred with each other. "Are you saying that multidimensional aliens attacked us?" one cried.

"Well … not exactly." She clapped her hands and the hologram vanished. "If not for this lucky recording, we'd never have known. You see, this wasn't an attack. Frankly, our extraterrestrial neighbours didn't know life occupied this side of the membrane at all."

"Dr Hennig, you're not making any sense —"

"What we saw," the scientist dared interrupt, "was an interdimensional weapon test."

Again, a deadly silence filled the chamber. "Explain," someone managed.

"I'm not sure how much I *can* explain. Like us, these aliens are involved in an arms race. Like us, they're always experimenting with bigger and better technology. Build a bomb, your enemies build a shield. So you build a better bomb. What we just witnessed …"

"Yes?"

"… was the test of an extradimensional missile. Put simply, it's a missile that winds up into a higher dimension — *our* dimension — and is then fired back into their dimension, in order to achieve extraordinary hypervelocities. Imagine having a fist-fight a few feet underwater. You're limited in how much power you can pack into a punch. But if you draw your fist through the air above …"



ILLUSTRATION BY JACEY

She flexed her gloved fingers again, replaying the holo to a molasses-like crawl. The freakish alien missile appeared, and was gone in a blink. Eris Station was pulled in after it, like debris sucked into an undertow.

"That's what happened to Eris Station," Hennig said decisively. "It got pulled into a neighbouring dimension along with the missile."

She noticed a new expression washing through the admiralty board. Excitement.

"Tell us more …" one said, practically licking his lips.

"Certainly, sir. The test resulted in the deaths of millions of our alien neighbours."

"No doubt. Looks like a badass missile."

"No, sir, you misunderstand. The missile detonation was fierce, yes, but it was a localized event. The alien test inadvertently caused a virulent epidemic that spread through half a continent." She regarded her audience's perplexed faces. "Because when the people of Eris Station were sucked into the interdimensional wake, their organic matter caused a deadly epidemic in the universe next door. Put simply, the aliens wanted a fast missile and ended up creating a bioweapon."

"Will this … happen again?"

"It's *been* happening, all over the Solar System. Brief flashes of light. Navigator reports of odd, transitory shapes glimpsed in space. But the aliens haven't been able to achieve the original results, because they aren't accidentally sweeping up human beings with their subsequent tests. Unfortunately, in their determination to recreate the effects, they've been increasing their tests. It's only a matter of time before a dimensionally spun-up missile materializes in the same space as a planet."

The admirals looked stricken.

"Don't worry," Hennig said.

"Don't worry?" an admiral mocked. "This stupid pleb has discovered a threat to the Imperium and she's telling us not to worry!"

"Perhaps she's found a way to neutralize this threat?" another suggested.

"Neutralize?" Hennig scratched her head. "Not exactly. You see, over the past few weeks I managed to establish contact with our interdimensional neighbours. Oh, it took a while, but we worked out a rudimentary form of communication. Surprised the hell out of them … but it's been productive, after a fashion."

"Wait, you initiated communication with a foreign power —"

"We made a deal," she spoke over the admiral. "They agreed to stop blindly testing their missiles in our dimension. In return …" She smiled. "We give them coordinates for where they *should* be testing their missiles."

She saw the realization kindle in their eyes. Or maybe, she thought, that was just the kindling of greenish light suddenly appearing around them.

The admiralty board leapt to their feet.

"We plebs are tired of being slaves to the Imperium," Hennig had to shout above the escalating sound. "So we sent the aliens coordinates of your military bases. The aliens get their bioweapons, and we get to be free of you. Everyone wins!"

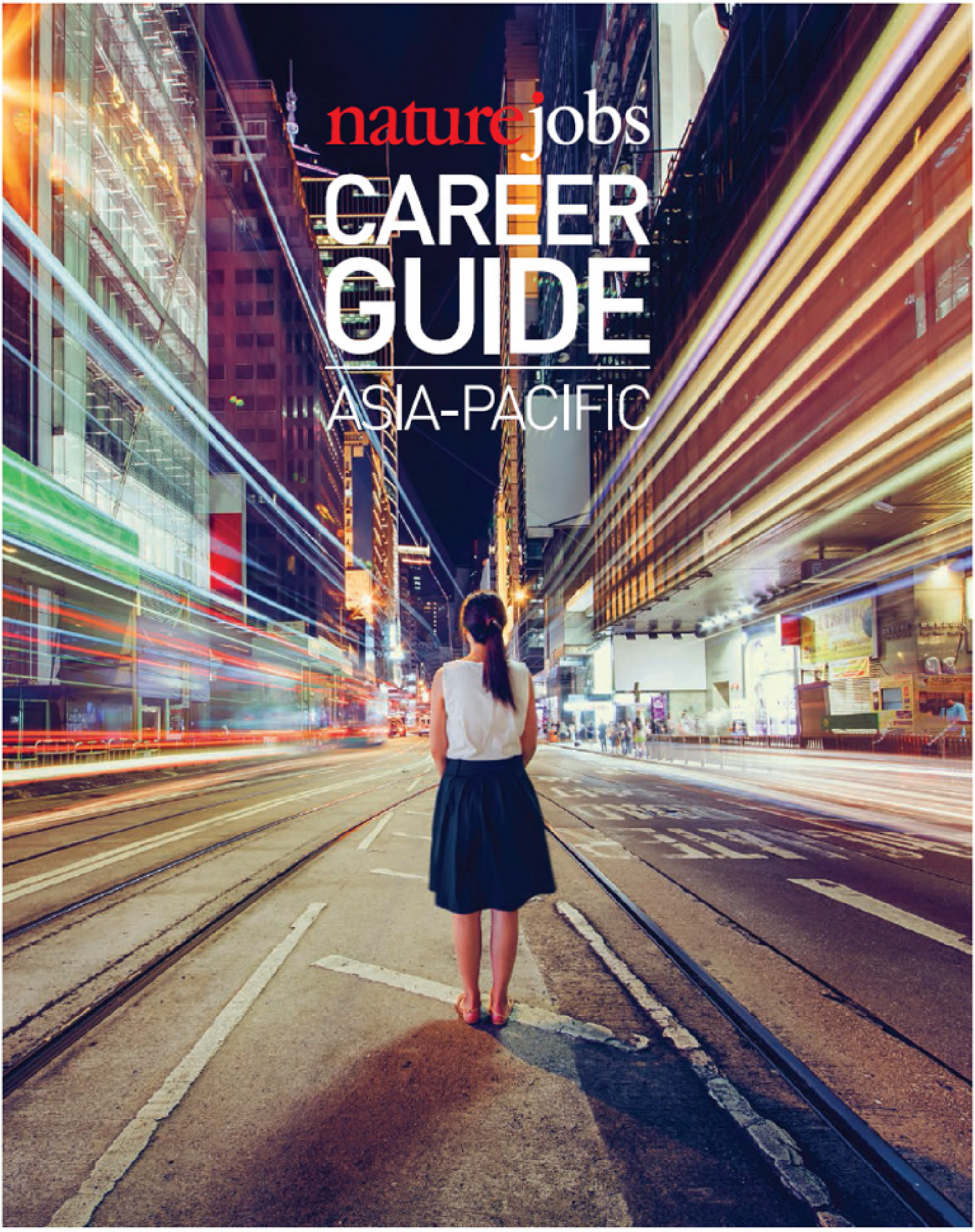
One of the admirals screamed in the intensifying light. "You'll die too!"

Hennig looked puzzled. "Me? Oh … I'm not even here."

She clapped her gloves, and her hologram vanished. ■

Brian Trent's *speculative fiction appears regularly in* ANALOG, Fantasy & Science Fiction, COSMOS, Galaxy's Edge *and numerous year's best anthologies. His website is www.briantrent.com.*

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



nature jobs

CAREER GUIDE

ASIA-PACIFIC

nature.com/careerguides/apac-2016

SPRINGER NATURE

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

naturejobs CAREER GUIDE ASIA-PACIFIC

NATURE, VOL. 536, NO. 7617 (25 AUGUST 2016)

COVER IMAGE: YU YU HOI/THE IMAGE BANK/GETTY IMAGES

Making a career decision can be like trying to solve an equation. Both tend to be easier when they involve fewer variables. Unfortunately, scientists who are thinking about where to work to achieve their goals must balance multiple priorities, often on the basis of poor quality data.

One thing that is clear is that as competition for research positions and funding grows in the West, opportunities are opening up in the East. China doubled the proportion of its gross domestic product that it spends on research and development (R&D) from 1% in 2002 to 2% in 2014 — and it plans to increase this to 2.5% by 2020. By then, the Organisation for Economic Co-operation and Development predicts that China will have overtaken the United States as the country that spends the most on R&D.

South Korea's investment of 4.3% of its GDP on R&D in 2014 was the largest of any country in the world. In January, Singapore unveiled a five-year plan to boost government support for R&D by 18%. Japan, while also increasing funding, is stepping up its efforts to attract international students and researchers by making more tuition available in languages other than Japanese.

Money isn't everything, of course. Australia continues to be a magnet for international researchers by focusing on established strengths such as nanoscience. New Zealand also concentrates on its specialities such as earth and environmental sciences. Both countries have the happiest people of the nations profiled in this supplement. New Zealand came eighth and Australia ninth in the United Nations *World Happiness Report 2016 Update*.

India is not yet a major player in science. Still, some of its leading institutions increasingly see international recruitment as key to the country achieving its scientific ambitions.

The 2016 *Naturejobs* Career Guide to the Asia-Pacific contains facts and figures on metrics such as research quality, the proportion of the workforce in research and development, salaries and cost of living. It also presents tips from employers and employees on how to get a job and how to deal with cultural differences in each of the seven featured countries.

The hope is that this information can help simplify the decision-making equations faced by those thinking about moving to the region, and in turn increase the chance of finding career solutions that meet their expectations.

Nic Fleming
Guest Editor

CONTENTS

S2 RELOCATING SCIENCE
A graphical guide to choosing the right country to work in

S6 CHINA
Foreign researchers are benefiting from high government investment

S14 JAPAN
Breaking down of language barriers is helping to attract researchers to the country

S18 AUSTRALIA
Quality of life and research reputation are drawing scientists down under

S22 SOUTH KOREA
Robotics is one of the many fields that has benefited from the country's investment

S26 INDIA
Embracing outside experience could help the country achieve its ambitions

S28 SINGAPORE
The city state is investing in research in the hope of producing economic returns

S30 NEW ZEALAND
A small science community can mean opportunities for some

EDITORIAL: Tom Benner, Herb Brody, Nic Fleming, Rebecca Kanthor, Smriti Mallapaty, Karen McGhee, Annabel McGilvray, T. V. Padma, Jenny Rooke, Mark Zastrow **ART & DESIGN**: Wesley Fernandes, Annthea Lewis, Alisdair Macdonald **PRODUCTION**: Sue Gray, Karl Smart, Ian Pope, Chandler Gibbons **MARKETING**: Claire Jones **SALES**: Janet Cen, Yuki Fujiwara, Helen Hill, Nicole Yu **DIRECTOR, GLOBAL CAREER SERVICES**: Nils Moeller **PROJECT MANAGER**: Anastasia Panoutsou **ART DIRECTOR**: Kelly Buckheit Krause **EDITORIAL DIRECTOR, PARTNERSHIP & CUSTOM MEDIA**: Stephen Pincock **PUBLISHER**: Richard Hughes.

NATUREJOBS CAREER GUIDE
The *Naturejobs* Career Guide Asia-Pacific 2016, a supplement to *Nature*, is produced by Nature Research, part of Springer Nature. *Naturejobs* (naturejobs.com) is the global jobs board and career resource for scientists.

Nature Editorial Offices
The Macmillan Building
4 Crinan Street,
London N1 9XW, UK
Tel: +44 (0)20 7833 4000
Fax: +44 (0)20 7843 4596/7

CUSTOMER SERVICES
To advertise with *Naturejobs*, please visit naturejobs.com/support or email naturejobs@nature.com
feedback@nature.com
Copyright © 2016 Macmillan Publishers Limited.
All rights reserved.

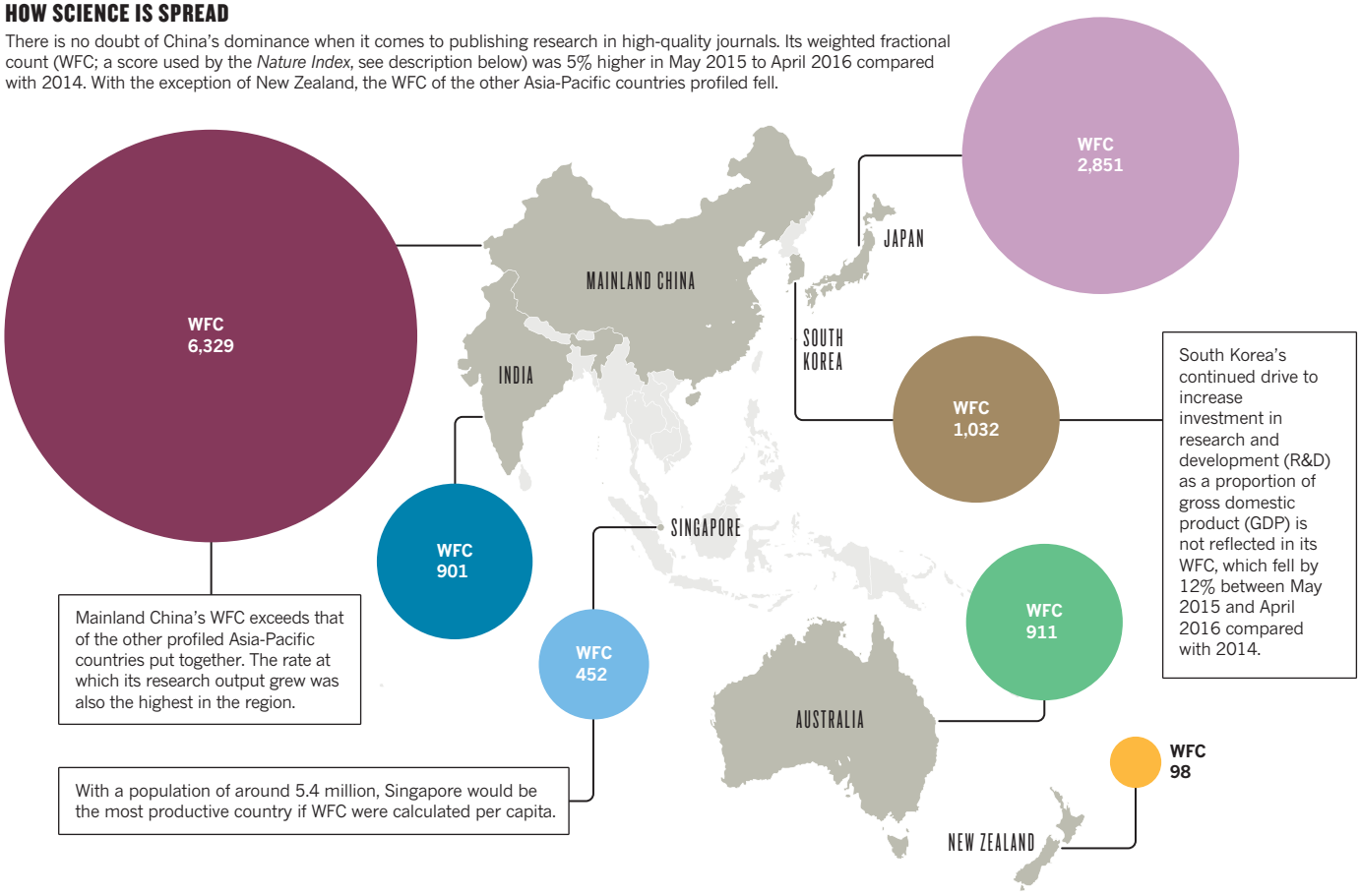
© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

RELOCATING SCIENCE

Countries are spending more than ever on research and development, but the fields they fund vary depending on national priorities. And it is not just the research reputation that matters when choosing whether to move abroad — cost of living and quality-of-life are factors too.

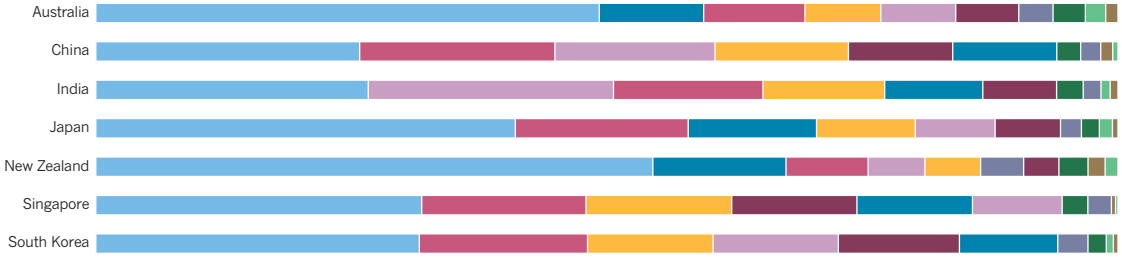
HOW SCIENCE IS SPREAD

There is no doubt of China's dominance when it comes to publishing research in high-quality journals. Its weighted fractional count (WFC; a score used by the *Nature Index*, see description below) was 5% higher in May 2015 to April 2016 compared with 2014. With the exception of New Zealand, the WFC of the other Asia-Pacific countries profiled fell.



SUBJECT STRENGTHS

China devoted more of its research efforts to physics and less to health and medical sciences in 2015 than any of the other countries profiled, according to the number of academic journal articles in the Scopus database. Articles on boundaries between subjects can be counted more than once, potentially distorting the relative subject proportions in these charts.



KEY

- Astronomy
- Biochemistry, molecular biology and genetics
- Chemistry
- Earth sciences
- Environmental sciences and energy
- Health and medical sciences
- Materials science
- Mathematics, computer science and engineering
- Microbiology and immunology
- Physics

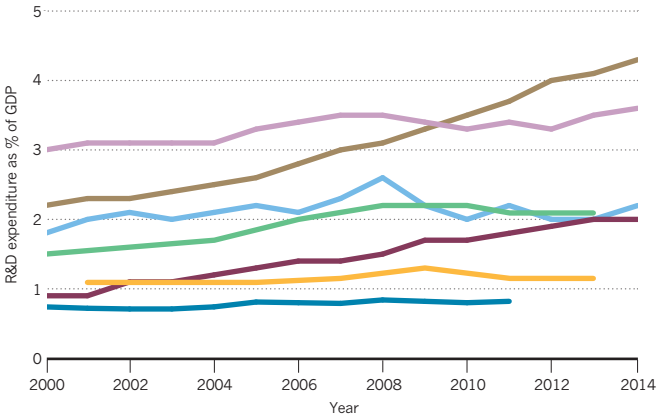
NATURE INDEX The *Nature Index* database tracks the affiliations of high-quality scientific articles, and charts publication productivity for institutions and countries. Weighted fractional count (WFC) accounts for the relative contribution of each author to an article and applies a weighting to correct imbalances in the index's subject coverage. This Career Guide draws on *Nature Index* data derived from articles published between 1 May 2015 and 30 April 2016. WFC is used throughout this supplement as the primary metric, because it provides an even basis for comparison. For more information, visit natureindex.com/faq

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

KEY Australia · China · India · Japan · New Zealand · Singapore · South Korea

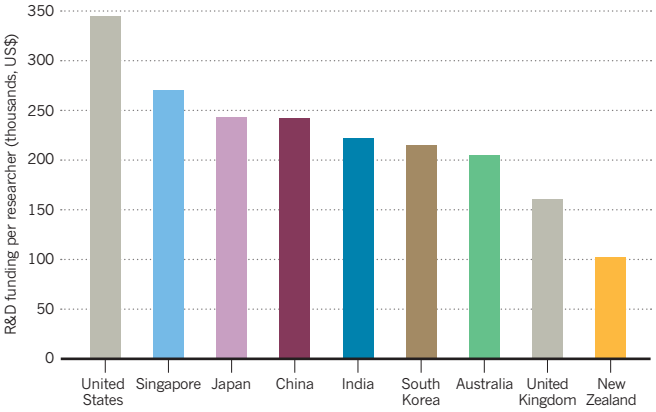
FUNDING OVER TIME

R&D funding as a proportion of GDP has grown most rapidly in South Korea and China since 2000, but has remained consistently low in India and New Zealand.



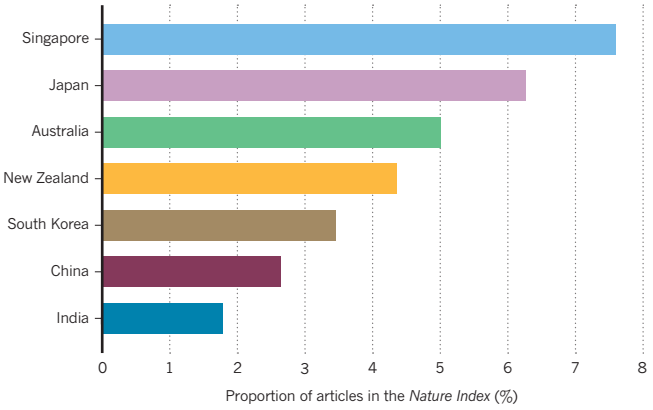
SPENDING PER RESEARCHER

Of the profiled countries, Singapore spends the most on R&D per researcher. Figures are the most recent available and are normalized for purchasing power. The United States and United Kingdom are included for comparison.



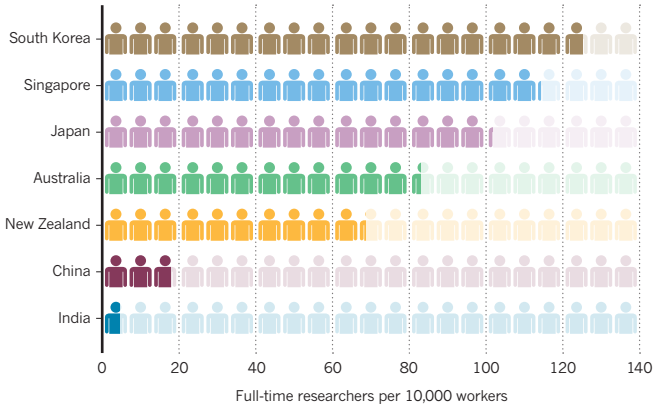
RESEARCH QUALITY

Singapore comes out on top for national research quality. Quality is based on the percentage of each country's biological and physical science articles in the Scopus database that make it into the *Nature Index*.



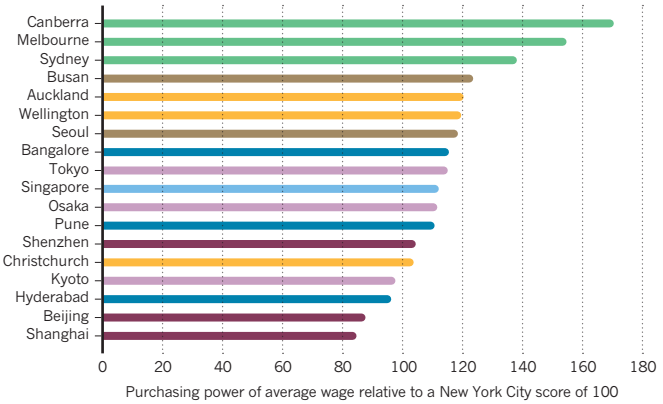
RESEARCHER DENSITY

South Korea has more of its workforce engaged in research than any of the other profiled countries. As a proportion of the working population, South Korea has around 30 times more researchers than India, according to the latest data available.



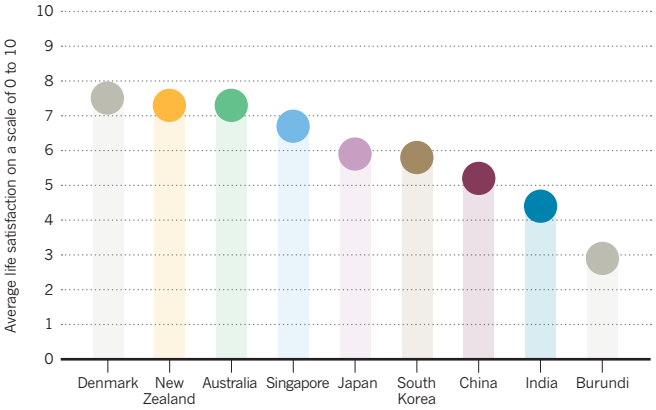
COST OF LIVING

Major cities in Australia are the most expensive places to live among the countries profiled, on the basis of an index of the relative ability of people on average salaries to buy goods and services.



LIFE SATISFACTION

When people of 157 countries were asked to rate life satisfaction on a scale of zero to ten, New Zealanders and Australians were the happiest people of the seven profiled countries. Denmark (the happiest) and Burundi (the least happy) are included for comparison.



SOURCES: FUNDING OVER TIME, OECD/UNESCO; SPENDING PER RESEARCHER, UNESCO; RESEARCHER DENSITY, UNESCO; COST OF LIVING, NUMBEO; LIFE SATISFACTION, J. F. HELLIWELL *ET AL.* IN *WORLD HAPPINESS REPORT 2016* (UN, 2016)

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



VCG/GETTY

China's Five-hundred-meter Aperture Spherical Telescope will help the search for extraterrestrial life.

CHINA

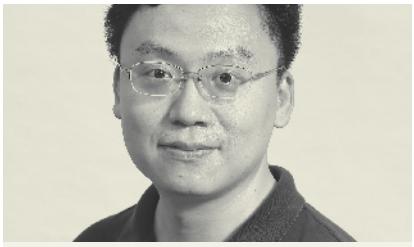
As China continues to increase its investment in research, it is offering opportunities that can be difficult to find elsewhere.

BY REBECCA KANTHOR

In 2007, while visiting his then girlfriend, US astronomer Eric Peng was working at a rented desk at Peking University in Beijing when he was approached by a fellow scientist. "You keep coming back," the man said. "Are you looking for a job here?" Six months after this casual conversation, Peng was offered a faculty position in Peking University's Kavli Institute for Astronomy and Astrophysics. "I think a lot of people thought it was an obvious thing to do, but I never thought I would actually move here."

China's economy has been growing more slowly in recent years: the gross domestic product (GDP) grew by 6.9% in 2015, compared with 10.6% in 2010. But foreign researchers say that there has been no cooling of the enthusiasm for enticing talented scientists to work in the country. Overseas and expat scientists are persuaded with promises of generous funding and research opportunities. China's most recent five-year plan, for example, pledged to continue with its drive to increase research and development funding as a proportion of GDP — from 1% in 2002 and 2% in 2014, to 2.5% in 2020.

The government launched its 1000 Talents plan in 2008 to attract scientists, entrepreneurs and finance experts from across the world. Precise figures for how many foreign scientists work in China are hard to come by, but according to anecdotal reports the incentives are working at least in some places. PAGE S8 ▶



JIANG GE
Associate provost at ShanghaiTech University

Why should foreign scientists come to China?
The economy is strong, so the conditions are good and money for research is very stable. In fact, the government is increasing investment in basic research, not only for large-scale facilities, but more generally. I see Chinese people coming back from working in other countries, as well as foreigners coming to work here. Although some have had good opportunities abroad, here they have excellent equipment and there is support for those starting up their own research.

What do Shanghai and ShanghaiTech University offer foreign researchers?
Shanghai is unique in being a huge city with such a high density of facilities. And there are more facilities planned for the future. Researchers want a good research environment, with good academic partners and the funds to purchase the materials they need. That is what is on offer at ShanghaiTech University. Shanghai is also home to many scientific centres at which researchers from all over the country work on projects to serve the national interest. These include the Shanghai Synchrotron Radiation Facility, the National Center for Protein Science and the Free Electron Laser Facility. These large-scale projects are funded by the central government and the Shanghai local government.

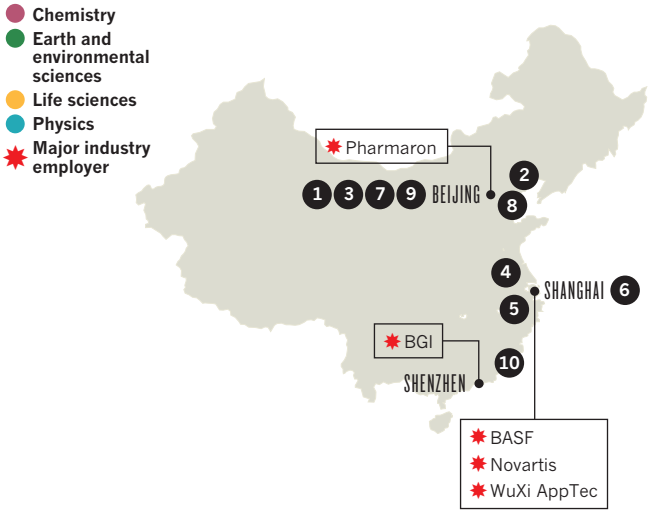
What skills do foreign researchers need to work in China?
We don't have enough top-quality researchers to help China to reach the level of research we are aiming for, so that's what we're looking for. Shanghai is an international city so it is not essential to speak Chinese — English is widely understood among the research community. **R.K.**

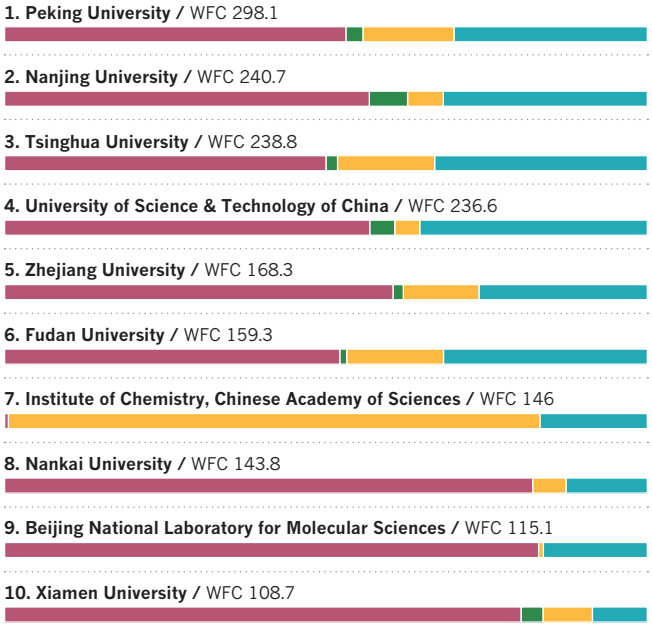
This interview has been edited for length and clarity.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

WHERE TO WORK

The top ten institutions in mainland China, based on research output included in the *Nature Index*, May 1 2015–April 30 2016, shown as weighted fractional count (WFC), a measure of the relative contribution of an author to an article weighted to correct for imbalances between subjects. Bars are divided according to the proportion that each subject area contributes to the overall score.

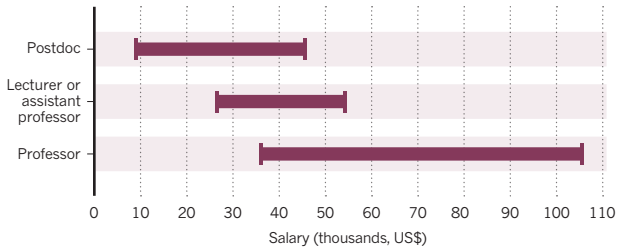




Overlaps in subject areas may cause some distortion to relative subject proportions.

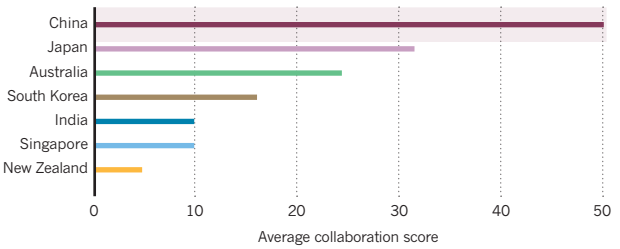
SALARIES

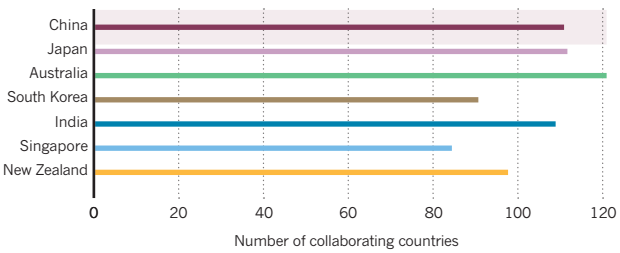
Chinese starting salaries are among the lowest in the Asia-Pacific countries profiled, especially for those lower down the pay scale, according to data collected in *Nature*'s interviews.



COLLABORATIONS

China's average collaboration score (top) — the sum of *Nature Index*'s fractional count (the relative contribution of authors to an article) for international collaborations divided by the number of countries China collaborates with.





RESEARCH FOCUS

The search for extraterrestrials gets a boost this year with the completion of China's new Five-hundred-meter Aperture Spherical Telescope (FAST). As well as hunting for life on other planets, the world's largest radio telescope, which was finished in July after 22 years of planning and 5 years of construction, will also allow astronomers to study pulsars — the cores of exploding stars — and survey hydrogen in faraway galaxies.

Built at a cost of 1.2 billion yuan (US$180 million) in a valley in the remote, mountainous region of the southern Guizhou province, FAST is made up of 4,450 triangular panels fitted into a giant dish with a surface area of 36 American football fields. Its 500-metre width smashes the previous record, held by the 305-metre wide Arecibo radio telescope in Puerto Rico. Some 9,000 people have reportedly been relocated to allow FAST to be built.

The enormous size of this instrument, which has been nicknamed 'Sky Eye', makes it more powerful and sensitive than previous radio telescopes and allows it to survey a larger part of the sky. Because it should be capable of picking up weaker radio signals from further away than any previous telescope, FAST will be particularly useful in the hunt for extraterrestrial life.

FAST forms part of China's plan to build up its space programme. Having put its first astronaut in space in 2003 and having launched its first unmanned lunar expedition in 2013, the country is planning to have an operational space station by around 2020 and to put a person on the Moon by 2030. ■



Staff at the construction site of FAST — the world's most powerful radio telescope.

OU DONGQU/XINHUA/EYEVINE

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Peng says that half of the postdocs and 20% of the faculty at his institute are from abroad.

The rapid growth of research funding in China makes it an attractive place for researchers looking to set up their own laboratories and projects. For Peng, it meant the opportunity to co-found the Telescope Access Program. Now in its sixth year, the US$13-million project gives Chinese astronomers observation time on optical and infrared telescopes around the world.

Ross Howie, a physicist who completed his PhD at the University of Edinburgh, UK, three years ago, is enjoying the greater autonomy his move to China in 2015 has allowed him. He is investigating materials synthesis under extreme pressures and temperatures at the Center for High Pressure Science and Technology Advanced Research in Shanghai. "I have been given a platform to build my own laboratory and have almost complete freedom to pursue my research goals," says Howie.

"I HAVE ALMOST COMPLETE FREEDOM TO PURSUE MY RESEARCH GOALS."

Despite such opportunities, relocating to China can come with challenges. Many grant proposals must be written and defended in Mandarin, although some organizations, such as the National Natural Science Foundation of China, have started to accept grant proposals in English. Scientists complain of excessive paperwork and frequently changing policies. Research institutions can reallocate grant money to other projects that don't have funding. And Internet censorship hinders research.

Jordanian neuroscientist Nashat Abumaria, at Fudan University in Shanghai, has seen such challenges defeat a number of foreign researchers since he arrived in the country in 2007. But he enjoys the drive and ambition of his Chinese colleagues, some of whom he has developed strong friendships with. Abumaria has turned down job offers from elsewhere and plans to stay as long as he can. "I've learned one important thing," he says. "You can not expect China to change for you. You must change and adapt with China."

Peng advises those willing to give working in China a try to think like an entrepreneur to make the most out of their experience. "There are definitely failings in the system," he says. "But there are a lot of advantages too." ■



CHINA HAS 50 UNESCO WORLD HERITAGE SITES, SECOND IN THE WORLD ONLY TO ITALY.

ENTRY REQUIREMENTS

- Employers must apply for work permits on behalf of employees. This includes presenting CVs, university transcripts and certificates, and making the case that no Chinese national can fulfil the job requirements.
- Once a work permit has been granted, employees must apply for a single-entry work visa called a Z visa.
- On arriving in China, the employee and their employer have one month to apply for a residence visa. The visa can take up to three months to process and is valid for one year, after which it must be renewed annually.
- The bureaucratic hurdles associated with visas and work permits can be frustrating; some have reported having to fly back to their home country to wait for documents.

OPPORTUNITIES & CONTACTS

- The 1000 Talent Plan of Foreign Experts offers attractive salary and research funding packages to those willing to work in China for three years.
- The 1000 Talent Plan for Young Professionals seeks to attract scientists aged under 40 who are seen as future leaders in their fields.
- The President's International Fellowship Initiative provides Chinese Academy of Sciences funding to PhD students and postdocs.
- The National Natural Science Foundation offers grants to both foreign and returning Chinese scientists seeking to collaborate with Chinese scientists.



ZAI LAB

SAMANTHA DU
Chief executive officer at Zai Laboratory, Shanghai

What is it like working in biotechnology in China?
The government wants to promote innovation, so it is very interested in the biotech sector. Both Chinese and global investors are keen to invest in it. Many Chinese entrepreneurs, people in the pharmaceutical industry and researchers want to come back to China to start their own businesses. Compared with the United States and other developed countries, Chinese biotech is still at an early stage, but it's catching up fast. The sector is booming.

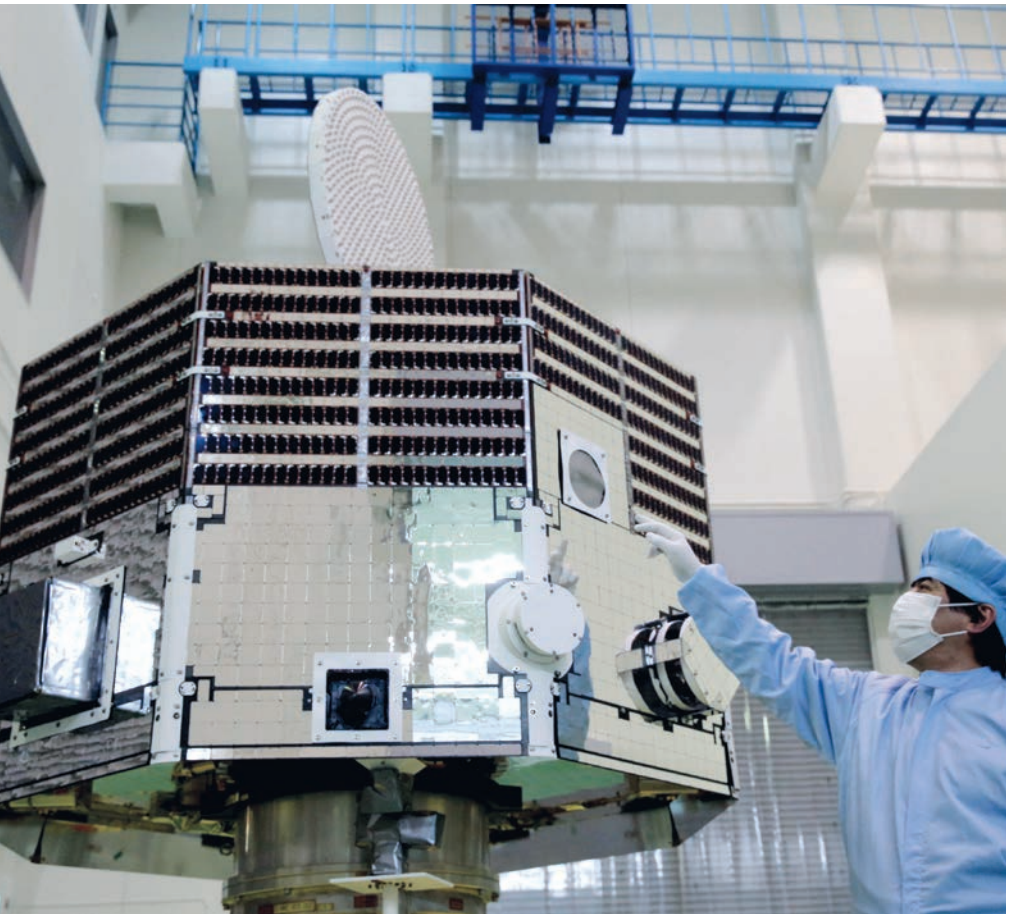
What skills are biotech companies looking for in foreign scientists?
They are the same as any US biotech company. I see a lot more opportunities opening up. We have recruited a group of overseas talent that has tremendous experience in drug discovery and development.

How will foreign biotech scientists benefit from working in China?
It will help them to understand biotech culture and pharmaceutical research in China. Eventually, if they want to go back to their home country they will have a lot of connections here. It also gives them the opportunity to return to China later, possibly as a representative for US or European companies.

What is the best thing about working in biotech in China?
It's very dynamic. For example, the government recently approved a pilot plan that will allow contract organizations to produce drugs. Those working in biotech in China are at the forefront not just of scientific discovery, but also of a generation pushing for new regulations that will pave the way for the development of truly innovative medicines in China. **R.K.**

This interview has been edited for length and clarity.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



THE ASAHI SHIMBUN/GETTY

The Mercury Magnetospheric Orbiter will study the planet's atmosphere and magnetic field.

JAPAN

The government is stepping up efforts to attract international scientists, as the country invests record sums in research.

SMRITI MALLAPATY

Cevayir Coban has witnessed a major cultural shift in Japanese universities since arriving in the country in 2003. Coban, her husband and their one-year-old son had left the United States so that she could join the laboratory of immunologist Shizuo Akira at Osaka University.

At first, Coban, a malaria immunologist from Turkey, kept herself to herself because she couldn't understand what her colleagues were saying. "The attitude back then was: 'If you don't speak Japanese, you should learn it'," she says.

That sense of isolation has eased, partly because she learned the language, but also because universities have changed their approach to non-Japanese speakers: they now seek out top-level international researchers to work in their labs. "In the past ten years, universities and institutions all over Japan have made an active effort to globalize," Coban says. "The attitude today is: 'You don't have to speak Japanese to work here'."

Around the time of Coban's arrival, the Japanese government, mired by a long period of economic stagnation, recognized that the country was ill-equipped to operate in the rapidly globalizing international research environment. In 2007, it set out its World Premier International Research Center Initiative (WPI), a plan to establish centres of excellence that would attract talented international researchers. Coban now heads her own lab at one of these centres: Osaka University's Immunology Frontier Research Center (IFReC), which is led by Akira. The centre is fully bilingual, 25% of its researchers are non-Japanese and it employs a team of staff to help researchers to relocate and deal with the inevitable paperwork involved in PAGE S16 ▶



CAROLINE FERN BENTON
Vice-president at the University of Tsukuba

What kind of researchers do you hire?
The university looks for individuals who are open to cross-disciplinary and cross-border collaboration. Being an open, global university is very much part of our DNA, and it is reflected in our philosophy and history. One of our prominent founding figures was Kanō Jigorō — the creator of modern judo and the first person to bring foreign exchange students to Japan.

In 2014, as part of a national effort to raise the international prominence of Japanese universities, we introduced an initiative to become a trans-border university in every sense of the word. We want to cross national, institutional, disciplinary and academic–industry borders. We are aiming to triple the number of foreign researchers at Tsukuba to more than 300 by 2023.

Why are Japanese universities so keen to globalize?
The number of 18-year-olds in Japan, and therefore the number of prospective university students, is declining. This means that we will need to bring in bright, young people to fill the university places and eventually to join the workforce. Research and education have also become much more global. We need to motivate our researchers to go abroad to expand their horizons.

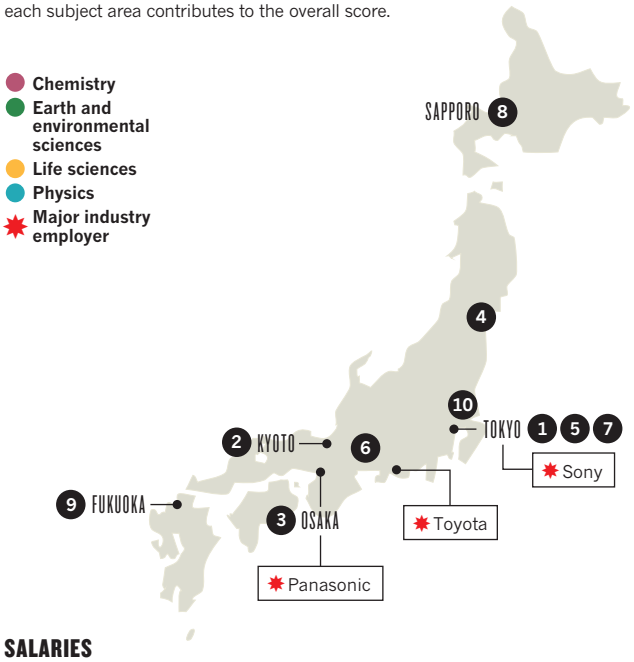
What do you enjoy most about working in Japan?
Japanese culture values harmony and teamwork. This means you feel a sense of belonging to a group. These values are promoted through relationships and a common understanding, so employers tend to take a longer-term view of employment and business. Even research labs at private companies tend to have a much longer-term focus than US companies, for example, which face greater pressure to deliver short-term results. **S.M.**

This interview has been edited for length and clarity.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

WHERE TO WORK

The top ten institutions in Japan, based on research output included in the 2015 *Nature Index*, May 1 2015–April 30 2016, shown as weighted fractional count (WFC), a measure of the relative contribution of an author to an article weighted to correct for imbalances between subjects. Bars are divided according to the proportion that each subject area contributes to the overall score.

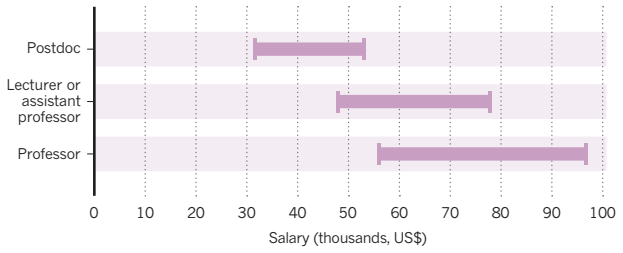


1. **University of Tokyo** / WFC 434.8
2. **Kyoto University** / WFC 259.8
3. **Osaka University** / WFC 180.1
4. **Tohoku University** / WFC 166.8
5. **RIKEN** / WFC 138.6
6. **Nagoya University** / WFC 126.9
7. **Tokyo Institute of Technology** / WFC 114.1
8. **Hokkaido University** / WFC 97.4
9. **Kyushu University** / WFC 83.5
10. **National Institute for Materials Science (NIMS)** / WFC 71.9

Overlaps in subject areas may cause some distortion to relative subject proportions.

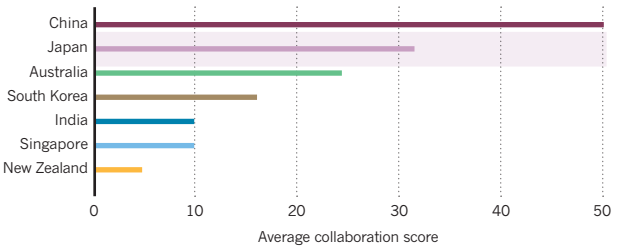
SALARIES

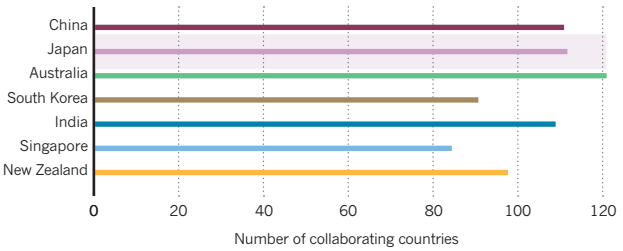
Starting salaries in Japan sit around the middle of the list of Asia-Pacific countries profiled, according to data collected in *Nature*'s interviews.



COLLABORATIONS

Japan's average collaboration score (top) — the sum of *Nature Index*'s fractional count (the relative contribution of authors to an article) for international collaborations divided by the number of countries Japan collaborates with.





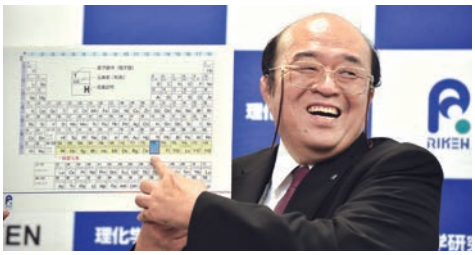
RESEARCH FOCUS

Japan will become the first Asian country to name a chemical element, it was announced in the final hours of 2015. An international committee of chemists had attributed the discovery of element 113 to a team led by chemist Kosuke Morita at the RIKEN Nishina Center for Accelerator-Based Science in Wako.

Elements with more than 92 protons are unstable and decay into lighter, more stable atoms after only a fraction of a second. Proving the existence of these ephemeral elements is therefore very difficult. Morita's team began its search for element 113 in 2003 by bombarding atoms of the heavy metal bismuth with beams of zinc travelling at one-tenth of the speed of light. It took the group 9 years to produce convincing evidence that it had created an isotope of an element with a nucleus of 113 protons and 165 neutrons, and another 3 years for the feat to be formally acknowledged.

The team has proposed the name nihonium, with the symbol Nh, after *Nihon*, Japanese for 'Japan'. The elements place, on the seventh row of the periodic table between copernicium and flerovium, is expected to be confirmed by the International Union of Pure and Applied Chemistry in November, following a five-month public consultation on the name.

In January, the government adopted a five-year plan that identifies robotics, sensor and actuator technologies, biotechnology, human–computer interaction, nanotechnology and quantum technology as areas of existing research strength that should be prioritized for investment. The government is also committed to funding a major cybersecurity project to safeguard crucial transportation, energy and communications infrastructure. ■



Chemist Kosuke Morita led the team that discovered element 113.

KAZUHIRO NOGI/AFP/GETTY

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

applying for grants and other administration.

In 2014, the Japanese government stepped up its efforts to make the country's higher education more competitive and compatible with the rest of the world by selecting 37 universities to receive funding under its Top Global University Project. The programme aims to increase the proportion of researchers at these institutions who are either foreign-born or Japanese with overseas degrees from 28% to 47%. Moreover, the aim is that by 2023, 22% of classes will be taught in languages other than Japanese — close to triple the current percentage. "We need to capture the power of globalization and openness

"THE ATTITUDE TODAY IS: 'YOU DON'T HAVE TO SPEAK JAPANESE TO WORK HERE'."

to address global and social challenges together," says economist Yuko Harayama, an executive member of the Japanese government's Council for Science, Technology and Innovation.

Despite these efforts, Japan is still falling short of its goals. The number of international researchers in appointments and placements of more than 30 days at Japanese academic and research institutions reached a peak of around 15,100 in 2012 — only a slight increase on the 12,800 researchers in 2003, and a small proportion of the 844,000 researchers in Japan.

"We have been promoting globalization for more than 20 years, but have not yet achieved the status we would like," says Harayama. Universities are under multiple pressures, she adds. They are required to open up to foreign influences, collaborate more with the private sector and focus to a greater degree on science that serves society.

Japan also falters when it comes to women in science. Not only is Coban one of the first international professors in the natural sciences at Osaka University, but she is also the only female principal investigator at IFReC out of almost 30. Nationally, women make up only about 15% of researchers and only 6% of corresponding authors — compared with an Organisation for Economic Co-operation and Development average for female corresponding authors of 26%.

But Harayama is optimistic about the future. In January, the government released a new science and technology plan, which was prepared with greater private-sector involvement than previous strategies — industry accounts for 72% of research and development expenditure in the country. Starting in April, organizations that employ more than 300 people are required to set minimum targets for employment of women and the proportion of women in management roles.

Meanwhile, Japan remains committed to investing in science and technology. In 2014, it spent a record high of 3.6% of its gross domestic product on research and development, making it the third highest investor globally after South Korea and Israel. ■



JAPAN HAS CLOSE TO 30,000 NATURAL HOT SPRINGS, OF WHICH ALMOST 8,000 HAVE BEEN DEVELOPED INTO POPULAR BATHING FACILITIES KNOWN AS ONSENS.

ENTRY REQUIREMENTS

- Researchers must apply for a Highly Skilled Foreign Professionals visa, which grants workers, and, if applicable, their spouse permission to live and work in Japan for up to five years.
- The application process takes between one and three months. Researchers must first obtain a certificate of eligibility from their employer. This is issued through a system that awards points based on criteria such as academic and professional experience. Employees include this certificate when applying for a visa at their Japanese embassy.
- On entering Japan, arrivals receive a residence card that they must carry at all times. Card holders can leave as many times as they wish, but they must re-enter the country within a year.
- Visa-holders can apply for permanent residency after five years.

OPPORTUNITIES & CONTACTS

- The Japan Research Career Information Network (JREC-IN) Portal lists jobs for researchers in academia, industry and the public sector.
- The Japan Aerospace Exploration Agency (JAXA) International Top Young Fellowship offers early-career researchers the opportunity to work at the Institute of Space and Astronautical Sciences for three to five years.
- The RIKEN Special Postdoctoral Researchers Program supports young researchers engaged in an independent project of their choosing for up to three years.
- The Japan Society for the Promotion of Science Postdoctoral Fellowship for Overseas Researchers offers grants to highly qualified researchers to collaborate with Japanese scientists for between one month and two years.



NTT BRL

BILL MUNRO

Group leader of the Theoretical Quantum Physics Research Group at NTT Basic Research Laboratories (BRL) in Atsugi

Are there any cultural differences that researchers coming to work in Japan should be aware of?

No one really says yes or no, so you need to get used to reading between the lines. If someone tells you that something 'will be difficult', then you probably shouldn't do it. But many Japanese people have now learned to be more direct with their non-Japanese co-workers.

How open is the private sector to international researchers?

Many industries are hiring more foreigners. About 15% of NTT BRL employees are non-Japanese and this proportion is slowly increasing. International researchers can apply for two-year postdocs. Older, grey-haired people like myself can apply for five-year research specialist positions.

Is private-sector research in Japan done differently?

Overseas, the trend is for industrial research to become more applied, more product-oriented, but in Japan most of the big companies still carry out fundamental research. Companies here also cover a wider breadth of research fields, and the labs are incredibly well-resourced.

Are there any downsides to working in Japan?

Historically, private-sector research has been a very individual activity in Japan. This is changing, but more slowly than in Western countries, where companies are moving towards larger, collaborative projects. A researcher can only do so much on their own. NTT laboratories are pushing towards more collaboration, both internally and externally. **S.M.**

This interview has been edited for length and clarity.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



DAVID HANNAH/LONELY PLANET/GETTY

Many people coming to work in Australia find its cities cosmopolitan and welcoming places to live.

AUSTRALIA

Scientists from across the world are attracted to the country, which competes internationally by focusing on its strengths.

BY KAREN MCGHEE

Australian research has delivered a series of major advances, from the bionic ear and Wi-Fi to the cervical-cancer vaccine and the black-box flight recorder. It has played crucial roles in the development of Google Maps, penicillin and ultrasound medical imaging.

The country's history of innovation and its model of scientific development are seen by many as consequences of its geography and demography. Australia's comparatively small population of 24 million lives in a vast country that is not only far removed from the world's major centres of development, but also dominated by an arid interior. It's of little surprise, then, that Australians have a reputation for resourcefulness and self-reliance.

Melbourne and Sydney, which are home to almost half the population, are the nation's hotbeds of scientific endeavour. Each has more than a dozen internationally ranked universities and research institutes that have assembled international teams in specialist fields.

"We eat lunch every day with German, American, French, Chinese, Iranian, Spanish, Singaporean and Australian researchers," says J. J. Richardson, a bioengineer at Australia's largest research body, the Commonwealth Scientific and Industrial Research Organisation (CSIRO) in Melbourne. Richardson, who moved to Australia from the United States in 2010, has been working alongside his Chinese colleague Kang Liang on biological applications of a class of crystalline materials called metal–organic frameworks.

Liang says that Australia has a reputation for making occasional, rather than PAGE S20 ▶



SWINBURNE UNIV. TECH.

LINDA KRISTJANSON
Vice-chancellor at Swinburne University of Technology in Melbourne

How does Australia's geographical distance from Europe and the United States shape its research?
Distance is not a barrier. We build research alliances and new opportunities for our researchers and students through our partnerships and shared quest for knowledge. For example, our astronomy students in Melbourne have access to the most powerful telescope in the world, at the Keck Observatory in Hawaii, through both remote access and site visits. This has allowed our astrophysics team to be ranked among the best in the world.

Our researchers routinely collaborate with other leading scientists and academic leaders from around the world through participating in conferences, collaborative research projects and joint initiatives. Our digital connections keep us engaged, and we think in a global way to ensure the work we do has impact.

And what about collaborations within the Asia-Pacific region?
Australia is exceptionally well located to work with our academic and research partners in the Asia-Pacific region. We can undertake research that addresses issues of mutual concern, such as energy, food security, manufacturing, climate change and caring for older people.

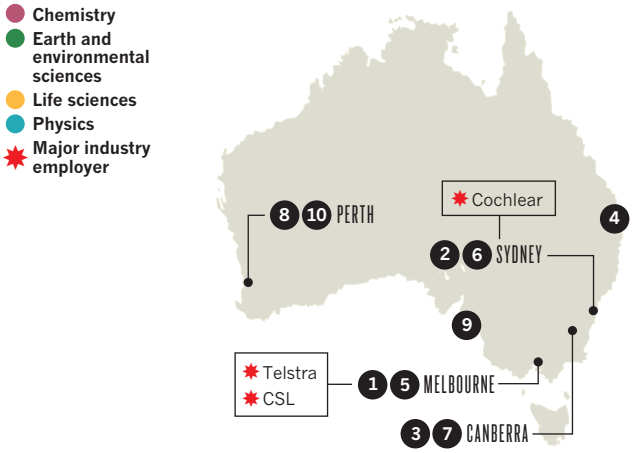
What is it about Australia that made you stay?
When I first moved to this country from Canada 20 years ago to pursue a career in palliative-care research, I found Australia to be a resilient country that welcomes people with aspirations and a readiness to engage and reach high. Quality of life, an intellectually stimulating environment, and people's willingness to embrace diversity, adopt new technologies and build innovation capability make Australia an attractive place to pursue a scientific career. **K.M.**

This interview has been edited for length and clarity.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

WHERE TO WORK

The top ten institutions in Australia, based on research output included in the 2015 *Nature Index*, May 1 2015–April 30 2016, shown as weighted fractional count (WFC), a measure of the relative contribution of an author to an article weighted to correct for imbalances between subjects. Bars are divided according to the proportion that each subject area contributes to the overall score.



1. Monash University / WFC 91.0

2. University of New South Wales / WFC 87.6

3. Australian National University / WFC 86.5

4. The University of Queensland / WFC 86.2

5. The University of Melbourne / WFC 82.4

6. The University of Sydney / WFC 73.1

7. The Commonwealth Scientific & Industrial Research Organisation / WFC 43.8

8. The University of Western Australia / WFC 38.3

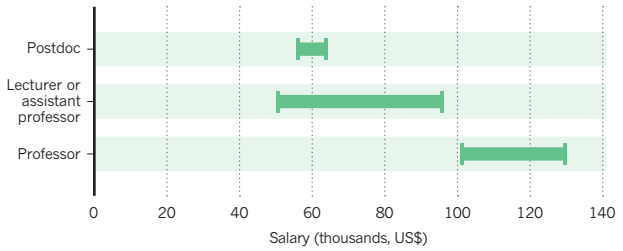
9. The University of Adelaide / WFC 31.5

10. Curtin University / WFC 22.2

Overlaps in subject areas may cause some distortion to relative subject proportions.

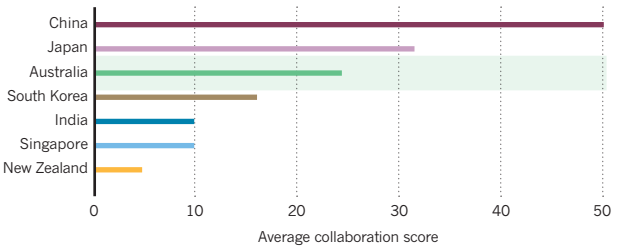
SALARIES

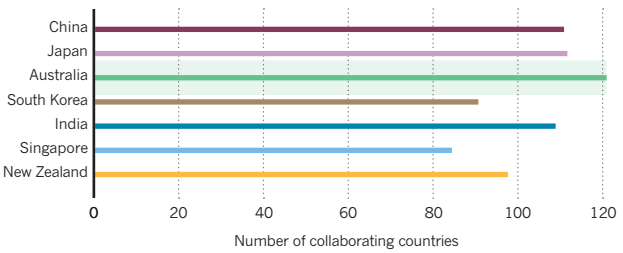
Postdocs in Australia are the highest paid in the Asia-Pacific countries profiled, according to data collected in *Nature*'s interviews.



COLLABORATIONS

Australia's average collaboration score (top) — the sum of *Nature Index*'s fractional count (the relative contribution of authors to an article) for international collaborations divided by the number of countries Australia collaborates with.





RESEARCH FOCUS

The opening of the Sydney Nanoscience Hub has provided a boost to Australia's already rising international reputation for nanoscience research.

Unveiled in April, the Aus$150 million (US$112 million) precision built and engineered facility took six years to design and construct. It is the centrepiece of the new Australian Institute for Nanoscale Science and Technology, based at the University of Sydney.

Nanoscience is the study of nanoscale materials (a nanometre is one billionth of a metre) and involves working with electrons, molecules, atoms and photons. Experiments can sometimes last only trillionths of a second. Because of this, research needs to be conducted in precisely controlled environments. The Sydney facility meets this demand by offering researchers laboratory space designed to minimize possible disturbances such as dust, pressure and temperature fluctuations, mechanical vibrations and electromagnetic radiation.

One of the Hub's three initial flagship projects focuses on optical physics and nanoscale photonics — hardly surprising when you consider that photonics (the study and application of light) is used in communications, imaging and data storage, underpinning a multi-trillion-dollar industry. One team, led by Benjamin Eggleton, the director of the Institute of Photonics and Optical Science at the University of Sydney, is attempting to develop a photonic chip that's faster and more energy-efficient at gathering and processing information than conventional electronic devices. ■



The University of Sydney is home to the new Sydney Nanoscience Hub.

KOKKAI NG/GETTY

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

consistent, scientific advances. "It's like a bubble that pops and a great innovation comes out, then it recedes for a while," he says.

Partly because of its small population, Australia has developed a number of specialisms, on which it focuses its efforts and resources. Health and medical research, for instance, has brought the country seven Nobel prizes. Physics and chemistry researchers have claimed three.

"AUSTRALIANS ARE NOT AFRAID OF TRYING NEW THINGS."

Astronomer Brian Schmidt shared the 2011 Nobel Prize in Physics for the discovery of the accelerating expansion of the Universe through observations of distant supernovae. US-born Schmidt, who is vice chancellor of the Australian National University in Canberra, is among many advocates who have spoken out about boosting the country's science, technology, engineering and mathematics (STEM) capacities.

There is certainly room for improvement. In 2012, for example, just 0.4% of those entering tertiary education studied mathematics, compared with the 1% average reported for countries in the Organisation for Economic Co-operation and Development. The lack of home-grown STEM talent provides opportunities for postdoctoral and PhD applicants from other countries.

Although overall research and development expenditure as a proportion of gross domestic product (GDP) has remained constant at a little over 2% in recent years, the funding of specific subjects has fluctuated as a result of changing political priorities. For example, 75 climate-science positions at CSIRO will be lost this year as a result of successive government cuts to the organization's budget in 2014 and 2015.

The geographical distance to international centres in the Northern Hemisphere can put some people off relocating to Australia. Sydney is around a 20-hour fight from either London or New York, for example. But many view Australia's remoteness as an advantage.

"Being a little distant from the mainstream, you're not influenced so much by what everyone else is doing; sometimes it means more space to pursue your own research interests," says Kelan Chen, a postdoc at the Walter and Eliza Hall Institute of Medical Research in Melbourne. Chen came to Melbourne from China to do her undergraduate degree and is now investigating an epigenetic regulator implicated in a form of muscular dystrophy.

A five-hour flight away in Perth, Sophie Monnier, a French PhD student studing exploration geophysics at the University of Western Australia, agrees. "Australians are not afraid of trying new things," she says, "because they are not hindered by historical cultural values". ■



AUSTRALIA'S ISOLATED LOCATION MEANS THAT MORE THAN 80% OF ITS TERRESTRIAL PLANTS AND ANIMALS — INCLUDING KANGAROOS, WOMBATS, KOALAS AND PLATYPUSES — ARE FOUND NOWHERE ELSE IN THE WORLD.



CSIRO

LARRY MARSHALL
Chief executive of the Commonwealth Scientific and Industrial Research Organisation (CSIRO) in Canberra

How do you think the way science is done in Australia could improve?
There is a lot of great science being done and investment in science is quite high. But we need to get better at ensuring that more of this research yields innovations that have social, environmental and economic impact.

What are the forces shaping the research landscape?
There is too much 'science push' rather than 'market pull'. The science is being pushed out from universities and from organizations such as CSIRO. The approach is 'This is what we have done. What can you do with it?' We need to recognize that our national challenges should be shaping our priorities. That will be a difficult transition, but it is one that this country needs to make.

In what areas is Australia looking to recruit expertise from overseas?
One focus is on digital technology and big data. We could achieve huge productivity gains in precision agriculture, and our health services are ripe for disruption. We also need to increase our ability to understand the impacts of 'blue-economy' developments, such as new fisheries. Other groups are looking at ways to obtain information in real time from living biological tissue by embedding sensors for diagnosis and intervention.

Does the Australian public value scientific research?
There is huge support for science in Australia because people recognize how much it has contributed to their day-to-day lives. Despite this, there is not much knowledge of the way that scientists work. We need to get better at explaining the scientific process in a way that relates to most people. **K.M.**

This interview has been edited for length and clarity.

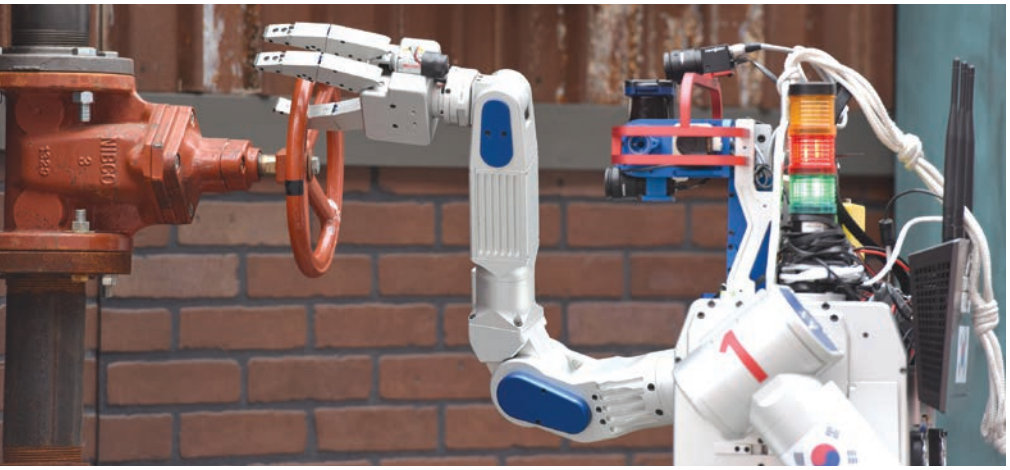
ENTRY REQUIREMENTS

- Most researchers coming to Australia use the Temporary Work (Skilled) 457 visa, which allows skilled workers with dependents to work in their chosen field for up to four years. Applicants may be sponsored by an employer or considered a good fit for an open position. The visa costs start at Aus$1,060 (US$800) and applications are usually processed within 90 days.
- Another option is the Skilled Independent 189 visa. This is aimed at English-speaking applicants under the age of 50 who are qualified to fill positions relevant to the country's skill shortage list, but who are not sponsored by an employer. The price starts at Aus$3,600 (US$2,587) and the visa is assessed under a points-based system.

OPPORTUNITIES & CONTACTS

- International Postgraduate Research Scholarships provide tuition fees and health cover for two years for master's degrees and three years for doctoral degrees.
- Commonwealth Scientific and Industrial Research Organisation (CSIRO) Postdoctoral Fellowships are for PhD graduates with no more than three years of postdoc experience.
- The National Health and Medical Research Council offers five-year fellowships for health and medical researchers.
- Forrest Research Foundation Scholarships are available to students who want to study towards a PhD at Curtin University, Edith Cowan University, Murdoch University, Notre Dame University or the University of Western Australia.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



DOD PHOTO/ALAMY

DRC-Hubo, the humanoid robot that won its creators from KAIST a US robotics prize.

SOUTH KOREA

A big investor in research and development, South Korea is attracting top scientists in the hope of boosting basic science.

BY MARK ZASTROW

When computer scientist François Rameau made the decision to pursue his career abroad, he did it more quickly than most. After looking up the research output of the robotics and computer-vision laboratory at the Korea Advanced Institute of Science and Technology (KAIST) in Daejeon, South Korea, where he had been offered a position, he instantly rejected the three postdoctoral fellowships on offer back home in France. "I realized it's probably one of the best in the world," he says. "I went without hesitation."

Rameau is now working on driverless cars — one of the hottest research fields in computer science. He has a five-year fellowship from the National Research Foundation (NRF) of Korea and additional funding from the German manufacturer Bosch. Since Rameau arrived, the lab's reputation has continued to grow: in 2015, a group of his colleagues won the US Defense Advanced Research Projects Agency Robotics Challenge with DRC-Hubo, a humanoid robot that can drive a car, climb stairs and open doors.

Robotics is one of the many fields that have benefitted from South Korea's focus on research and development (R&D) since the start of the century. In 2014, the country spent 4.3% of its gross domestic product (GDP) on R&D — the highest globally, and twice what it spent in 1999.

This investment has been partly spent on recruiting high-level international scientists to South Korean institutions. But these have often been joint appointments, under which the researchers work at the institutions for only a few months of the year. Although this approach may have lent prestige, it is widely viewed to have failed to improve research quality. As a result, many universities are now moving towards offering more long term, full-time appointments.

The Institute for Basic Science (IBS), a network of university-based research centres established in 2011, does not offer joint appointments. "Foreign recruits have to resign from their previous position and move here," says Doochul Kim, president of IBS, headquartered in Daejeon.

Efforts to attract talented scientists to South Korea are partly driven by concern over the country's low birth rate, and fears of a looming shortfall in the number of students taking up university places. The government hopes to double the number of foreign students at South Korean universities to 200,000 by 2023. It has also created a visa for entrepreneurs and offers citizenship to those who studied in Korea and have since been in employment for two years.

Language can present hurdles for foreign researchers — calls for grant proposals are often only in Korean, for example. And there are few international schools outside of the capital. This is why, instead of living close to his lab at KAIST, Turkish chemist Cafer Yavuz, and his wife and son, live in Seoul. He makes the three-hour commute by bus to KAIST twice a week. "It's a working solution, but it's not ideal," he says.

Rameau, who doesn't have children, has no such worries. He is considering looking for an academic position in South Korea when his fellowship ends. He says those moving to the country might find that it takes a while to get their bearings. "Don't judge too early," he says. "You need to stay for more than six months to start to understand the state of mind of the people." ■



KOREA INSTITUTE OF SCIENCE AND TECHNOLOGY

BYUNG GWON LEE
President of the Korea Institute of Science and Technology (KIST)

How easy is it to get a job at a South Korean institution?
The job market is getting tougher, but the country's increasing globalization offers great opportunities for foreign scientists.

What advice do you have for scientists hoping to work in South Korea?
Without learning Korean, foreign scientists cannot compete for R&D funding. For those who find work, I strongly advise that they broaden their professional and personal network. Tight social and business circles can limit opportunities, but Koreans are warm-hearted, and Korean scientists in particular are open-minded to scientists from abroad.

What are the relative benefits of working for government institutes, universities or industry?
Industry is well funded, but focuses on the profit motive to the extent that your project is continuously under review and can be terminated quickly. Government institutions also provide good resources, but it can be difficult to find high-quality students to fill positions, which can make it difficult to make progress. Academic institutions have access to very good students, but funding is sometimes limited.

How does South Korea's famously hierarchical work culture affect scientists?
Local researchers are very respectful towards principal investigators and senior researchers. Foreign researchers have a bit more freedom. They can freely walk out of the lab, whereas Koreans typically ask permission if they want to leave before the principal investigator.

What is the best thing about working in South Korea?
It's the Koreans. They are highly educated, and have strong characters and great energy. **M.Z.**

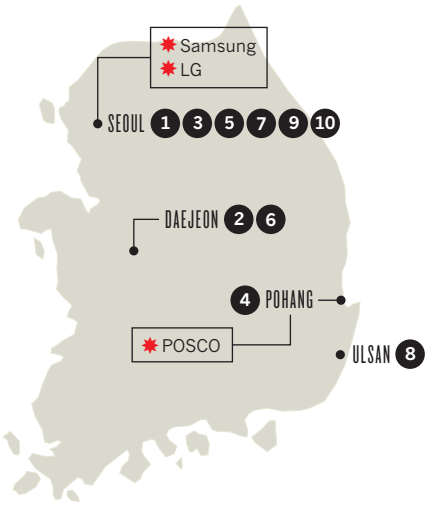
This interview has been edited for length and clarity.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

WHERE TO WORK

The top ten institutions in South Korea, based on research output included in the 2015 *Nature Index*, May 1 2015–April 30 2016, shown as weighted fractional count (WFC), a measure of the relative contribution of an author to an article weighted to correct for imbalances between subjects. Bars are divided according to the proportion that each subject area contributes to the overall score.

- Chemistry
- Earth and environmental sciences
- Life sciences
- Physics
- Major industry employer



1. Seoul National University / WFC 122.9

2. Korea Advanced Institute of Science & Technology / WFC 99.4

3. Yonsei University / WFC 80.7

4. Pohang University of Science & Technology / WFC 74.5

5. Sungkyunkwan University / WFC 54.1

6. Institute for Basic Science / WFC 52.1

7. Korea University / WFC 43.0

8. Ulsan National Institute of Science & Technology / WFC 37.2

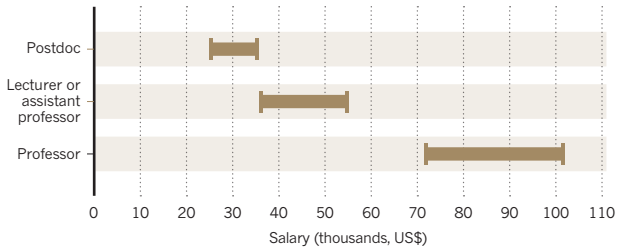
9. Hanyang University / WFC 32.1

10. Ewha Womans University / WFC 28.1

Overlaps in subject areas may cause some distortion to relative subject proportions.

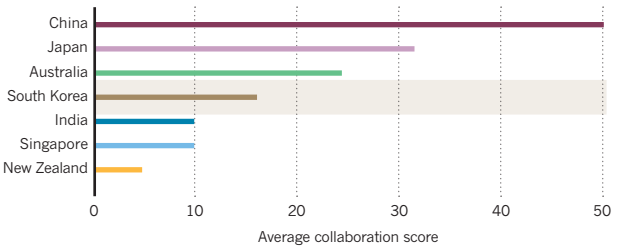
SALARIES

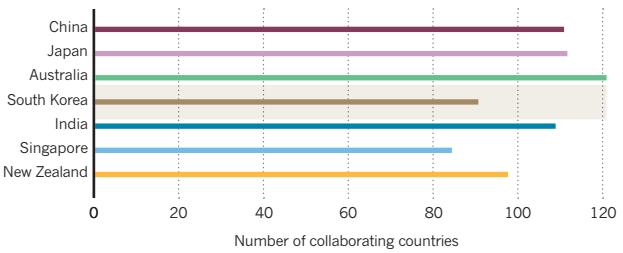
Compared with researchers elsewhere in the Asia-Pacific region, postdocs are relatively poorly paid and professors relatively well paid in South Korea, according to data collected in *Nature*'s interviews.



COLLABORATIONS

South Korea's average collaboration score (top) — the sum of *Nature Index*'s fractional count (the relative contribution of authors to an article) for international collaborations divided by the number of countries South Korea collaborates with.





RESEARCH FOCUS

The hunt for dark matter — which is thought to make up around 27% of the Universe — could come to an end in South Korea.

One possible component of dark matter is a theoretical particle called the axion. In 2017, physicists at the Center for Axion and Precision Physics Research (CAPP) in Daejeon expect to complete the development of a detector that could find this particle. CAPP director Yannis Semertzidis also hopes to build an experiment to detect the electric dipole moment of the proton, which if found could help to explain why there is more matter than antimatter in the Universe.

CAPP is one of 26 research centres that make up the Institute for Basic Science (IBS), a national institution founded in 2011 to bolster blue-sky research. The institute's task is to rebalance the country's scientific efforts to include more basic science alongside the previous focus on applied research, which was deemed to offer more immediate economic benefits. Other centres in the network focus on different fields, including laser science, gene editing and nanomedicine.

The IBS has plans to open further research centres that specialize in areas of mathematics, environmental science, theoretical and nuclear physics, optics, new materials and the use of nanotechnology in biomedicine. It plans to recruit scientists to serve as directors and associate directors of these, and offer fellowships for young scientists, in 2017.

OPPORTUNITIES & CONTACTS

- The Korean Government Scholarship Program offers scholarships to international graduates or undergraduates, consisting of one year of Korean instruction, followed by up to six years of study.
- The International R&D Academy at Korea Institute of Science and Technology trains scientists and engineers from developing countries and offers PhD programmes through the affiliated University of Science and Technology.
- The government-run Brain Korea 21 Program for Leading Universities and Students (BK21 PLUS) initiative funds professor and postdoc positions at various research universities. ■

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



XINHUA/EYEVINE

The Indian Space Research Organisation launch a rocket carring a navigational satellite.

INDIA

Although not a major scientific player, India hopes that attracting foreign researchers will help it achieve its ambitions.

BY T.V. PADMA

"India is a dynamic place with a fast-developing research landscape," says biophysicist Darius Koster. "Coming here provides an opportunity to bring in your own ideas and be part of this change."

Originally from Germany, Koster moved to India in 2011 to study the mechanisms of cell-membrane organization as a postdoc at the National Centre for Biological Sciences (NCBS) in Bangalore. As a foreign scientist working in the country, he is part of a small minority: in the past, Indian institutions have not tried to attract researchers from abroad.

India is not a major global scientific player. It has few scientists relative to its population, and many of its most talented researchers move abroad to work. The country invests less than 0.9% of its gross domestic product in research and development. Many think it needs to embrace experience from outside the country to achieve its ambitions. "We can no longer afford to have a frog-in-the-well attitude," says astrophysicist Tarun Souradeep at the Inter-University Centre for Astronomy and Astrophysics (IUCAA) in Pune. "Foreign scientists bring a diversity of research cultures and values."

A few institutions have begun to hire foreign postdocs and faculty members. These include the Tata Institute of Fundamental Research in Mumbai, the Indian Institutes of Technology, the Indian Institutes of Science Education and Research, the IUCAA and the NCBS.

India has big ambitions. In 2015, the government announced a plan to become a leading nation in terms of computing power by creating a network of around 70 supercomputers to connect the nation's academic and research institutions. It also revealed a five-year strategy to turn the country into a global genomics hub. And in 2018, the Indian Space Research Organisation plans to launch Chandrayaan-2, its second lunar mission, consisting of an orbiter, lander and rover.

The low price of goods and services relative to average salaries in India mean that researchers can enjoy a good standard of living. The average salary in both Pune and Bangalore buys more goods and services of an equivalent standard than the average wage in New York City, according to the crowdsourced database Numbeo. However, publicly-funded institutions do not contribute to pension schemes for foreign scientists, as they do for local employees.

Life in India can be challenging. "We knew that India would be hot, but nothing prepared us for our first summer here," says physicist Richard Morris, whose 2015 move from the United Kingdom with his wife and two daughters to work at the NCBS came shortly before a severe heat wave and drought. Axel Brockmann, a German honeybee specialist at NCBS who took an assistant-professor position in 2012, found it difficult to get used to the cows that roam freely in the road and to different attitudes to time-keeping in some quarters. "Here, 5 minutes can mean anything from 15 minutes upwards," he jokes. ■



NATIONAL CENTRE FOR BIOLOGICAL SCIENCES

SATYAJIT MAYOR
Director of the National Centre for Biological Sciences (NCBS) in Bangalore

What are the attractions of doing research in India?
India does not have many postdocs, so there is the potential, in this small system, for pursuing one's own vision. At a time when research resources are plateauing in most parts of the world, India has an opportunity to invest more — much as China and Japan are doing.

Are there any downsides for researchers coming to the country?
The notorious Indian bureaucracy is the country's Achilles heel. For foreign scientists, getting permission to work in India can be complicated. The slow delivery of funds from funding agencies and the creaking research infrastructure can also be an issue. And high-quality research is mostly confined to a few top institutes.

How has the NCBS attracted scientists?
The NCBS has an open structure, without rigid departmental boundaries. It works on biology at every scale — from molecules to the ecosystem — and has a strong interdisciplinary, multidimensional research culture. A good example is its programme on chemical ecology, which is the study of chemicals involved in the interactions within and between organisms. It is a unique opportunity to study uncharted territory in biology.

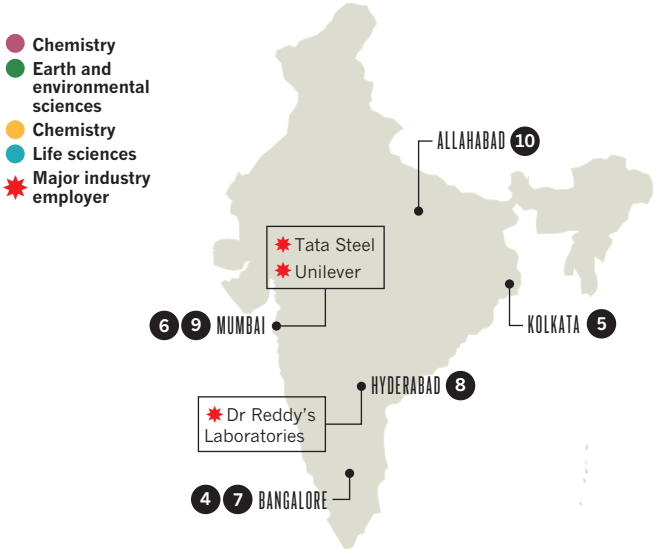
What opportunities does India offer?
As India's small scientific community grows, there is enormous scope to chart new directions in research across all disciplines. The Indian subcontinent has unique geological systems that reflect rare changes in Earth's geological history. It also has a diverse human genetic pool and is teeming with unexplored ecological niches — all of which offer great research opportunities. **T.V.P.**

This interview has been edited for length and clarity.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

WHERE TO WORK

The top ten institutions in India, based on research output included in the 2015 *Nature Index*, May 1 2015–April 30 2016, shown as weighted fractional count (WFC), a measure of the relative contribution of an author to an article weighted to correct for imbalances between subjects. Bars are divided according to the proportion that each subject area contributes to the overall score.

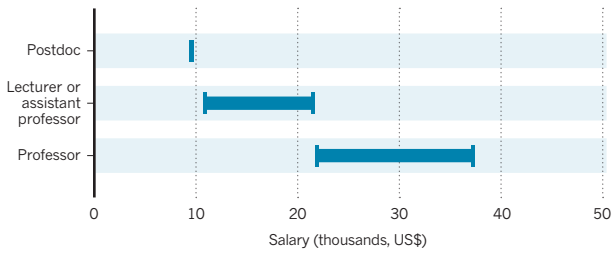


1. **Indian Institutes of Technology,** multiple locations / WFC 160.1
2. **Council of Scientific & Industrial Research,** multiple locations / WFC 131.3
3. **Indian Institutes of Science Education & Research,** multiple locations / WFC 89.4
4. **Indian Institute of Science** / WFC 76.4
5. **Indian Association for the Cultivation of Science** / WFC 43.7
6. **Tata Institute of Fundamental Research** / WFC 39.8
7. **Jawaharlal Nehru Centre for Advanced Scientific Research** / WFC 32.7
8. **University of Hyderabad** / WFC 21.6
9. **Bhabha Atomic Research Centre** / WFC 17.4
10. **Harish-Chandra Research Institute** / WFC 13.4

Overlaps in subject areas may cause some distortion to relative subject proportions.

SALARIES

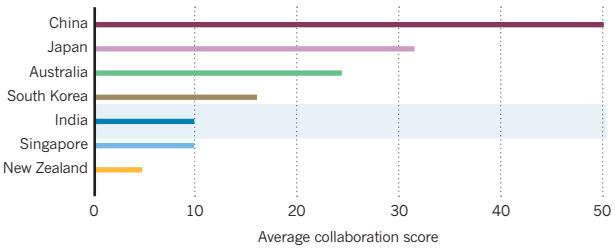
Starting salaries in India are the lowest of those in the Asia-Pacific countries profiled, according to data collected in *Nature*'s interviews*.

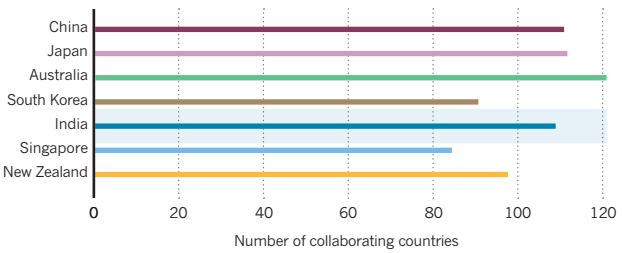


*Figures include a recently agreed pay rise for professors and lecturers or assistant professors.

COLLABORATIONS

India's average collaboration score (top) — the sum of *Nature Index*'s fractional count (the relative contribution of authors to an article) for international collaborations divided by the number of countries India collaborates with.





RESEARCH FOCUS

Ripples in space-time called gravitational waves were detected for the first time in 2015 — around 100 years after Albert Einstein predicted them in his general theory of relativity. Astronomers hope that the observations, made by twin Laser Interferometer Gravitational-Wave Observatory (LIGO) detectors in the United States, will provide a better understanding of gravity and answer other major questions about the Universe.

In February, India signalled its growing ambitions in astronomy by agreeing to build a third LIGO observatory, costing US$183 million. Alongside its US counterparts, LIGO-India will allow astronomers to pinpoint the source of the cataclysmic cosmic events that cause gravitational waves — such as the collisions of massive stars or black-hole mergers — much more accurately. The Inter-University Centre for Astronomy and Astrophysics has been advertising for postdocs and faculty members to support this research.

In 2015, India formally joined an international collaboration seeking to build the Square Kilometre Array (SKA), which would be the world's largest radio telescope, at sites in Australia and South Africa. It is leading the part of the consortium dedicated to the development of the hardware and software needed to control SKA. India has also emerged as a possible contender to host the Thirty Metre Telescope, the world's most advanced optical and near infrared telescope, following protests over the original plan to build on land in Hawaii considered sacred by native people.

OPPORTUNITIES & CONTACTS

- The Wellcome Trust/DBT India Alliance — jointly funded by the UK Wellcome Trust and India's Department of Biotechnology — offers Early Career Fellowships for postdoctoral biomedical researchers who want to develop their research careers in India. Intermediate and senior fellowships are also available.
- The Human Frontier Science Program, based in Strasbourg, France, offers postdoctoral fellowships to those who want to pursue basic life-sciences research abroad, including in India, in a field other than their own. ■

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



EDGAR SU/REUTERS

Singaporeans test a self-driving vehicle, technology that may soon be part of the transport system.

SINGAPORE

The government is spending more than ever before on research and development, especially on work that is likely to produce returns.

BY TOM BENNER

A global economic slowdown might seem like an unlikely driver of professional advancement for a scientist. Yet that is exactly what the financial crisis of 2007–08 was for French immunologist Florent Ginhoux. He began to consider working in Asia when the crash made it difficult to find an institution in the West that was willing to fund his plans to set up his own lab. In 2009, he moved his family to Singapore, where he became a principal investigator. He now leads a lab of 12 researchers and students, and has no intention of leaving any time soon.

"In Europe or the US, sometimes when you start your lab you have to think small and slowly expand," says Ginhoux, who is studying the biology of dendritic cells with his group at the Singapore Immunology Network (SIgN), a research institute that is part of the Agency for Science, Technology and Research (A*STAR). "Singapore was the only place where I could really think about doing the science with all the resources and support I needed."

The island city-state welcomes foreign talent such as Ginhoux as part of the government's commitment to promote research, innovation and enterprise as the cornerstones of its economy. A wealthy country, Singapore spends 2.2% of its gross domestic product on research and development (R&D). Its per capita R&D spending is among the highest in the world. In January, the Singaporean government unveiled a 5-year, S$19-billion (US$13.9-billion) plan to support R&D in the country — an 18% increase on 2011–15.

Up to S$4 billion of this will be invested in research collaborations between industry and academia, reflecting a broader trend in the way science is done in Singapore. In October 2015, for example, US-based engineering company Applied Materials established a joint R&D lab with A*STAR to develop advanced semiconductors for logic and memory chips.

Although government funding is available for basic research, an increasing amount is being directed towards work that is seen as most likely to have an economic impact. "There are still significant resources available for basic research at A*STAR, but there is a tangible push towards work that has a more immediate return on investment," says French geneticist Bruno Reversade, who joined Singapore's Institute of Medical Biology in 2008. He now leads a group that investigates the genetics of rare diseases and how twins are produced.

Renting property in Singapore is notoriously expensive, but the average salary there buys more goods and services of an equivalent standard than the average wage in London, according to crowdsourced website Numbeo.

Compared with neighbouring countries, western researchers may find Singapore easier to settle in. English is widely used, and there are plenty of international associations and schools for expats and their families. Ginhoux needed to do little to prepare for his move: "Singapore is open and foreigner-friendly," he says. ■



YALE-NUS COLLEGE

STEVEN BERNASEK
Dean of Faculty at Yale-NUS College

What has surprised you most about working in Singapore?
The biggest surprise is how quickly the landscape can change, and not just physically, in terms of buildings and infrastructure. Decisions to support particular areas of science or technology development are translated into financial support and research infrastructure much more quickly than they are in the United States or Europe.

What are the challenges and opportunities for foreign scientists?
Keeping up with the pace of change can be difficult, as can the need to rely on international suppliers for laboratory equipment and supplies. But many researchers are excited about doing science in an environment less constrained by resources than they may have been used to. There are also opportunities to work in areas of fundamental science that have clear connections to problems in society.

Should foreigners be aware of any differences in working cultures?
Singaporean researchers can sometimes be more deferential and less willing to challenge the ideas and plans of their team leader than those from elsewhere. There can be a tendency to work strictly to a job description, rather than thinking about ways to actually accomplish tasks and suggesting alternative ways to go about things.

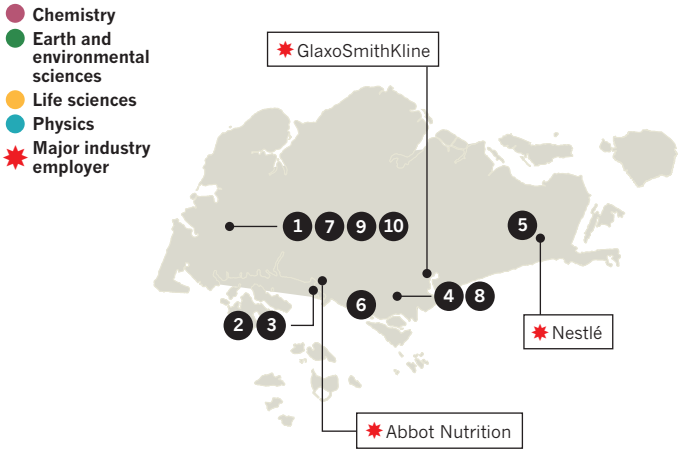
What is the best thing about working in the country?
It's the optimism about science and the support for science. It's such a contrast to doing basic science in the United States. **T.B.**

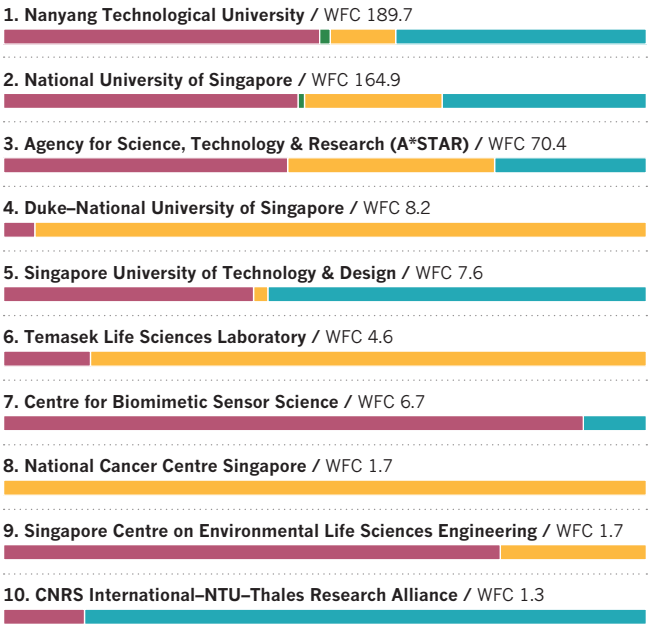
This interview has been edited for length and clarity.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

WHERE TO WORK

The top ten institutions in Singapore, based on research output included in the 2015 *Nature Index*, May 1 2015–April 30 2016, shown as weighted fractional count (WFC), a measure of the relative contribution of an author to an article weighted to correct for imbalances between subjects. Bars are divided according to the proportion that each subject area contributes to the overall score.

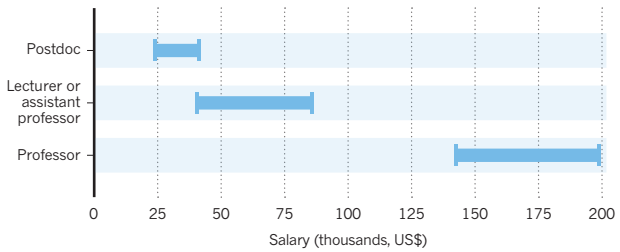




Overlaps in subject areas may cause some distortion to relative subject proportions.

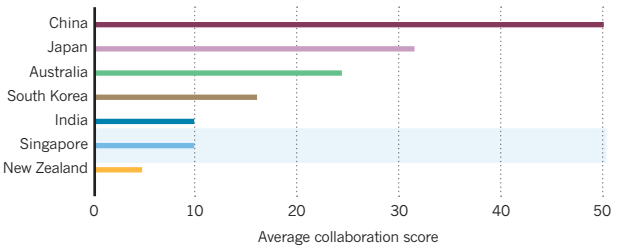
SALARIES

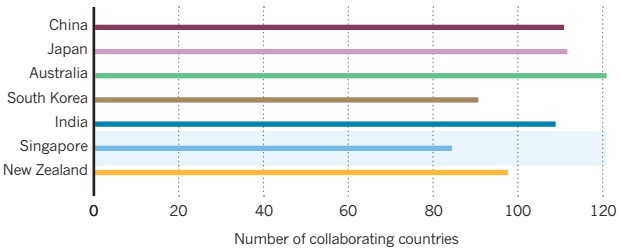
Starting salaries in Singapore are the highest among the Asia-Pacific countries profiled for professors, but less generous for those lower down the scale, according to data collected in *Nature*'s interviews.



COLLABORATIONS

Singapore's average collaboration score (top) — the sum of *Nature Index*'s fractional count (the relative contribution of authors to an article) for international collaborations divided by the number of countries Singapore collaborates with.





RESEARCH FOCUS

Imagine cars that know where empty parking spaces are; buildings that can identify sources of air pollution; and smartphone apps that alert users when someone nearby has a heart attack, and locates the closest defibrillator. Trend watchers have long predicted the advent of networks of connected sensors embedded in the physical objects around us, seamlessly collecting and exchanging data to enable a wide variety of innovations.

Many cities are experimenting with such technologies to improve traffic management or street lighting, for example. The Singaporean government's Smart Nation initiative, launched in 2014, outlines a broader, bolder vision of the Internet of Things. One early example is Beeline, an app that analyses historical travel patterns and crowdsourced data to suggest routes to bus operators. Smart Nation is creating systems that combine data from sensors such as those in smartphones, other sensors and surveillance cameras to generate a live picture of the nation in unprecedented detail. The initiative's fellowship programme offers funding to computer and data scientists with ideas for applications that would allow for better management of traffic congestion, power outages, floods and infectious-disease outbreaks, for example. The Singaporean government hopes that the country's compact size and centralized political system will allow it to become a world leader in the field. But given its history of authoritarianism, some are sceptical about the government's promise to anonymize the new data it will collect.

OPPORTUNITIES & CONTACTS

- The A*STAR International Fellowship is a two-year fellowship for postdocs who have completed their PhD in the past four years and final-year PhD students.
- The Lee Kuan Yew Postdoctoral Fellowship is a three-year fellowship at Nanyang Technological University (NTU) and the National University of Singapore, to be held concurrently with a staff research position. It is open to foreign students.
- NTU assistant professorships are tenure-track positions with attractive remuneration and start-up grant packages for foreign researchers who have obtained a PhD in the past ten years. ■

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



PLANT & FOOD RESEARCH

Jason Johnston works on automated monitoring of stored fresh produce at Plant and Food Research.

NEW ZEALAND

A small science community offers opportunities in a dramatic landscape, but can also limit career progression.

BY ANNABEL MCGILVRAY

When mathematician Alys Clark left the United Kingdom to do her master's degree in Australia, she only intended to be gone for a year. She ended up following her course with a PhD, before moving "across the ditch" to take up a job in New Zealand. Seven years after leaving home, she's still there, living with her New Zealand husband and their young son amid ferns and kauri trees, just a short drive from the black sands of the country's west-coast beaches.

"It's pretty amazing, in that you can drive for half an hour and you are on the beach and not far away are the mountains," says Clark, who uses mathematics and physics to create virtual organs and identify early pathological changes at the University of Auckland. In 2014, Clark was awarded a five-year Rutherford Discovery Fellowship, which she is now using to mathematically model the physical processes involved in early pregnancy.

Beyond the natural beauty on her doorstep, Clark feels at home in what she describes as a collegial, informal and inclusive research culture. "The senior academics are very approachable and supportive of researchers going off and setting up their own projects. The whole way of life is a little more relaxed than back home in the UK."

New Zealand has a population of less than 5 million and spends just 1.2% of its gross domestic product on research and development — half the international average, according to the Organisation for Economic Co-operation and Development. As a result, the science community is small and the scope for career progression is limited, particularly in academia. "It's not always an easy environment to stay in if you want a long-term career," says Clark.

But for others, the small pool has advantages. Having previously worked in his native Germany, as well as Switzerland and the United Kingdom, meteorologist Olaf Morgenstern moved to New Zealand to continue his climate-modelling research in one of the field's most challenging environments. Before the move, a colleague told him that whereas in larger communities he may be a small fish in a big pond, in New Zealand, "the pond is all yours". Indeed, since arriving to take a position at the National Institute of Water and Atmospheric Research in 2008, Morgenstern has become something of a big fish: last year he took on a job in Wellington, leading the Earth system modelling and prediction programme for the Deep South National Science Challenge, the country's Antarctic climate-science endeavour.

A growing interest in climate science and modelling in New Zealand in recent years fits well with the country's conventional strengths in Earth and environmental sciences. In the Asia-Pacific region, New Zealand leads the field, according to the *Nature Index*, which assesses research performance on the basis of contributions to high-quality publications. Agricultural research is a particular strength, driven by both government-owned research institutes and large corporations, such as the multinational dairy cooperative Fonterra. Although New Zealand's dramatic landscape is central to its appeal to many international researchers, for others, such as Morgenstern, the country's beauty is just an attractive backdrop to the scientific challenges and opportunities that it provides. ■



PLANT & FOOD RESEARCH

BRUCE CAMPBELL
Chief operating officer at Plant and Food Research, headquartered in Auckland

What aspects of working life in New Zealand make it appealing?
You can pursue an international-standard career, while also having a good work–life balance. There's recognition of the importance of family and a collegial work environment.

What distinguishes the country's research culture?
Plant and Food Research is one of seven government-owned Crown Research Institutes. Organizations such as ours have strong links with universities and industry groups, with secondments happening both ways, providing opportunities for diverse experiences. Because we are a small country, there are good opportunities for researchers to connect to key decision-makers in government and influence science-related policies. Science in New Zealand is currently undergoing reform to make the system much more collaborative between stakeholders such as research institutions, industry and government.

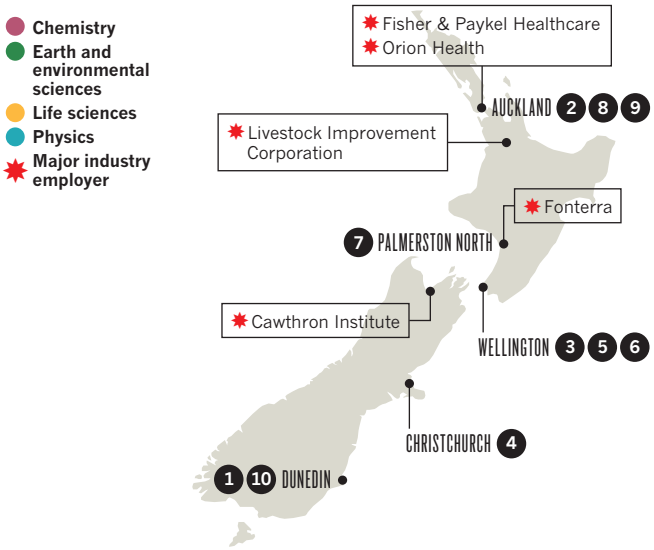
Does New Zealand's location make international collaborations difficult?
When you're working with researchers in other countries, there can be a perception that you're a long way away. We work hard to break down that barrier with modern digital communications. And scientists here probably spend more time on aeroplanes than those in other parts of the world. But it's only an overnight flight to the western United States or around 24 hours to Europe. Isolation can also stimulate innovation, and the physical environment in New Zealand is conducive to thinking about different ways of doing things, which is a great starting point for science. **A.M.**

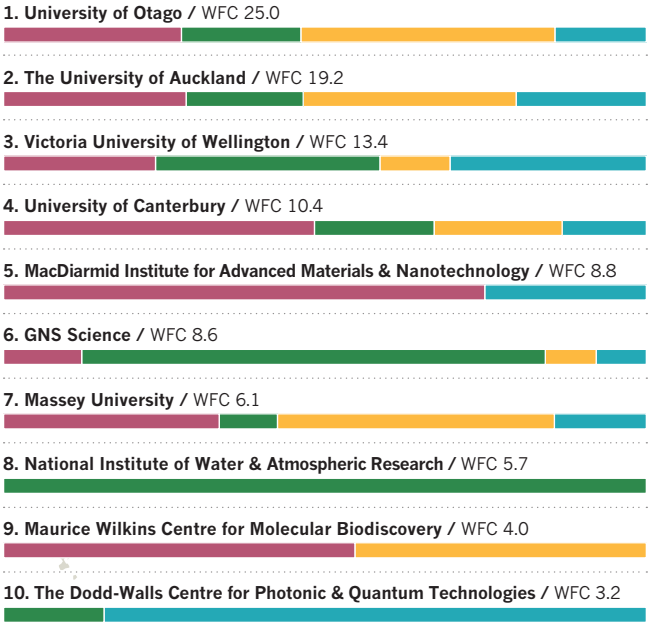
This interview has been edited for length and clarity.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

WHERE TO WORK

The top ten institutions in New Zealand, based on research output included in the 2015 *Nature Index*, May 1 2015–April 30 2016, shown as weighted fractional count (WFC), a measure of the relative contribution of an author to an article weighted to correct for imbalances between subjects. Bars are divided according to the proportion that each subject area contributes to the overall score.

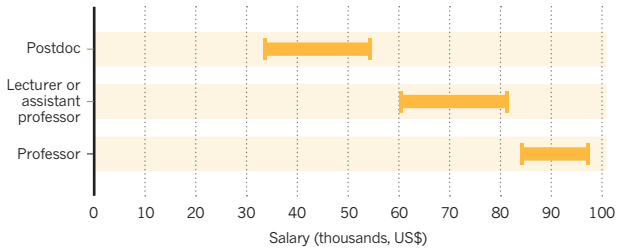




Overlaps in subject areas may cause some distortion to relative subject proportions.

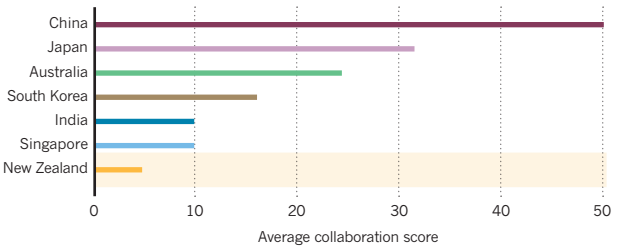
SALARIES

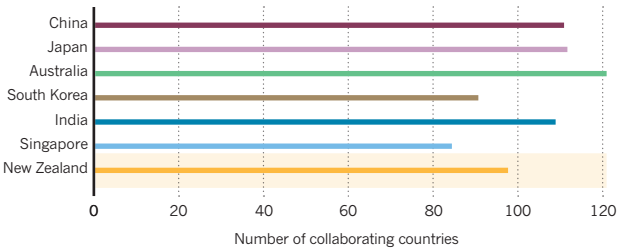
The relative gap between the best and worst paid in New Zealand is the smallest of those in the Asia-Pacific countries profiled, according to data collected in *Nature*'s interviews.



COLLABORATIONS

New Zealand's average collaboration score (top) — the sum of *Nature Index*'s fractional count (the relative contribution of authors to an article) for international collaborations divided by the number of countries New Zealand collaborates with.





RESEARCH FOCUS

Perched on the edge of the Southern Ocean, New Zealand is an ideal base from which to assess the roles that the Antarctic and its surrounding waters have in climate change.

Motivated by the most recent report by the Intergovernmental Panel on Climate Change — which says that existing international climate models do not accurately account for the unusual cloud-cover patterns caused by the low-pressure systems that dominate in the Antarctic — the New Zealand government has established the Deep South National Science Challenge. The multidisciplinary collaboration's mission is to better measure the changes occurring now, and more reliably predict what might happen in the future.

Deep South researchers are creating their own modelling system based on ongoing observations, and sharing their data with international groups to improve global modelling. Their results will also inform policies about domestic issues that could be affected by climate change.

Deep South is 1 of 11 National Science Challenges established by the New Zealand government to address issues of national and international significance. The others include Ageing Well, Sustainable Seas, and Building Better Homes, Towns, and Cities. The aim is to bring together publicly funded Crown Research Institutes, non-government bodies and business to tackle these challenges. The government is backing the plan with NZ$1.6 billion (US$1.2 billion) over a decade.

OPPORTUNITIES & CONTACTS

- The government's New Zealand Now service offers personalized e-mails about jobs opportunities, including those in research, as well as information on studying, working and living in the country.
- New Zealand International Doctoral Research Scholarships provide university tuition fees, living expenses and medical insurance for three years for international students undertaking a PhD in the country.
- The Li Lairong Horticultural Research Fellowship is available to Chinese students for research placements of up to three-month with Plant and Food Research, a government-owned institute. ■

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.





Aier School of Ophthalmology

Open positions at ASO

爱尔眼科
AIER EYE HOSPITAL

The Aier School of Ophthalmology (ASO), located in Changsha, the capital city of central China's Hunan province, was established by Central South University and the Aier Eye Hospital Group (Aier), China's largest eye care service provider. As the first specialized ophthalmic college in the country, ASO will focus on Aier's long-term development strategy of basic and clinical research, and leverage its strong financial position and rich clinical resources from over 120 affiliated professional eye hospitals, to train the next generation of talented ophthalmic researchers and professionals.

ASO is now inviting applications for the following positions:

1. Associate Dean of ASO
2. Vice Director of the Aier Research Institute of Ophthalmology
3. Research Principal Investigators (PI) in the fields of stem cells and precision medicine

Qualifications:

- A doctoral degree in medical science
- Prior employment as an associate professor or more senior level
- Demonstrated strong scientific research ability, with more than 10 publications in leading, peer-reviewed international journals
- A track record of securing external funding in overseas countries
- Excellent communication and organizational skills as well as a strong sense of responsibility

Salaries and benefits:

- Annual salary: 500,000-1,000,000 RMB (around 80,000-150,000 USD)
- Research funds as needed
- 2-3 research assistants (including PhD students)
- Accommodation for a transition period after relocation
- An offer of Aier share options
- An offer of Aier partnership program options
- Reimbursement of travel expenses for domestic and international academic meetings
- Incentives for publications in high-level journals
- Incentives for science and technology achievements
- Reimbursement of overseas travel expenses for up to two family-reunion trips a year
- Basic insurance and other fringe benefits in accordance with national regulations
- Recommendation and support for applications to various kinds of talent programs and funds

To apply:

Please send a curriculum vitae and a brief cover letter stating your career and research interests to Shibo Tang, professor and dean of ASO, Central South University at tangshibo@vip.163.com.

Beijing Advanced Innovation Center for Brain Disorders (BAIC-BD) Recruiting Principal Investigators

China Capital Medical University (CCMU) has the strongest comprehensive advantages in neuroscience in China and is the owner of the largest clinical resource of major brain diseases, epidemiology and epidemiological bio-bank. CCMU affiliated hospitals, such as Xuanwu Hospital, Tiantan Hospital and Sanbo Brain Hospital, specializing in major brain diseases are considered first rate in China and highly recognized in the world. CCMU and its affiliated hospital receive the nation's largest amount of patients with brain diseases, especially brain related hard-to-treat illness. CCMU equipped with abundant and solid research resources and platforms for neuroscience development, operates one National Clinical Disease Center for neurological and mental diseases, three National Key Disciplines, five National Key Specialties, two national centers for quality control, one Key Laboratory at provincial level, one Key Laboratory sponsored by the Chinese Ministry of Education and 11 Beijing municipal government sponsored Key Laboratories and Engineering Research Centers.

The Beijing Institute for Brain Disorders (BIBD) at CCMU, funded by the Beijing Municipal Government, is a research center aimed at understanding the mechanisms underlying brain disorders, and at translating results from the laboratory bench to the patient's bedside, in order to reduce the burdens that the disorders impose on patients, families, and society. As the preponderant subject carrier of CCMU, BIBD carry out the overall planning layout of the neuroscience discipline, discipline construction, personnel training and serve as the incubator of high level research & development. To achieve the major discoveries or inventions, form the platform for the scientific research, production tansforming and personnel training, BIBD performs the dual track running mechanism through the combination of public institutions and enterprises.



Beijing Advanced Innovation Center for Brain Disorders (BAIC-BD), established in 2015, is one of the first Advanced Innovation Centers sponsored by the Beijing municipal government. Hosted by the Capital Medical University and cooperated closely with BIBD, this center aims to advance the knowledge of brain diseases, translate science to treatment, and reduce the burdens of brain disorders. BAIC-BD strives to produce influential research discoveries and important healthcare products, establishing a world-class research center for the translational study of human brain diseases.

BAIC-BD welcomes dedicated and talented scholars to join from China and abroad. With a well-developed support system, the center provides a modern and scholarly environment for innovative research and will offer incentives to encourage creativity and innovation.

Responsibilities

1. Carry out research and participate in all types of research activities, including building up and supervising research teams, designing and conducting experiments, publishing high quality research articles, filing patents, etc.
2. Advance a specific mental health discipline, such as stroke, Alzheimer's disease, Parkinson's disease, brain tumor, depression, epilepsy, nerve injury and repair, and schizophrenia.
3. Mentor junior faculty members, supervise graduate students, and advise professional and undergraduate students.

Qualifications

Applicants should have a Ph.D. or M.D. in Neuroscience or related disciplines, be able to develop independent research, with over 5 year oversea experience, and have at least held associate professorship (or equivalent) in world-renowned universities or research institutes. Priorities will be given to those with over 10 years' experience in a relevant field of brain diseases and with high accomplishments.

Application procedure

Applicants should submit a curriculum vitae, 3-5 recommendation letters, copies of academic credentials and representative publications to liuyujia@ccmu.edu.cn and hwzheng@ccmu.edu.cn .

Further Contact

Address: Capital Medical University, Xitoutiao 10#, Youanmenwai, Fengtai District, Beijing 100069, China.
Contact Person: Ms Liu, Ms Zheng
Phone: +86-1083911208, +86-1083911894
E-mail: liuyujia@ccmu.edu.cn hwzheng@ccmu.edu.cn



Talents Wanted Programme

Faculty search for the School of Environment, Beijing Normal University

The School of Environment at Beijing Normal University is seeking faculty candidates under its newly launched and well-funded Talents Wanted Programme. All research fields covered by the School of Environment are eligible, including aquatic environments and ecologies, urban ecology, wetland habitats, basin environmental modification and ecological restoration, environmental engineering, hydrology and water resources, and environmental evaluation and management.

Qualifications:

Applicants should hold a PhD degree and are expected to be in their thirties or forties. Overseas postdoctoral experience will be considered an advantage. Applicants should have a demonstrated commitment to excellence in teaching and research at a level comparable to an assistant or associate professor in a top-tier international university. Successful candidates will be employed on a full-time basis and are expected to establish an internationally competitive, independent and cutting-edge research programme in a relevant field at the School of Environment.

Benefits of working with us:

- Successful candidates will be employed as principal investigators and will be able to supervise doctoral students.
- The School of Environment will offer an internationally competitive salary and assist successful candidates in finding accommodation on or around campus.
- The School of Environment will provide office and laboratory space as well as internationally competitive research start-up packages.

Application process:

Qualified applicants are strongly encouraged to submit their applications electronically to Ms. Jun Zhang (zj0207@bnu.edu.cn).
Applications should include the following materials (in PDF format):

- a comprehensive curriculum vitae;
- a research plan and teaching plan (maximum two pages);
- three to five references with contact information for the referees.

For further information please contact:
Ms. Jun Zhang
Tel: +86 10 58800397; Fax: +86 10 58800397
E-mail: zj0207@bnu.edu.cn
Website: http://env.bnu.edu.cn

北京師範大學環境學院
School of Environment, Beijing Normal University

DukeNUS Medical School

LEADING RESEARCH AND INNOVATION IN SINGAPORE AND ASIA

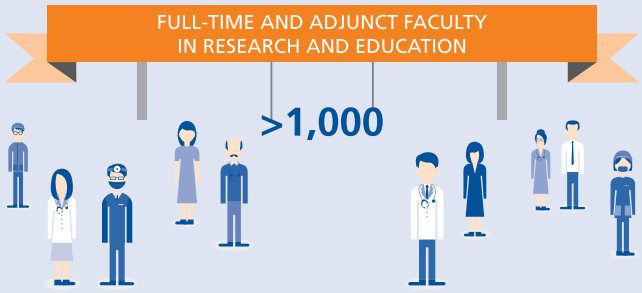
Backed by world-renowned Duke University and the National University of Singapore, Duke-NUS Medical School is a young, vibrant institution that punches above its weight.

Duke-NUS' international reputation in biomedical research and medical education has grown exponentially since 2005. Furthermore, its partnership with the collective clinical strengths of Singapore's largest healthcare provider, SingHealth creates a dynamic scholarly community for new discoveries, learning and care innovations which ultimately benefit patients.

SingHealth DukeNUS ACADEMIC MEDICAL CENTRE

With over S$350 million clinched in research funding, numerous research awards won and more than 2,800 papers published, our numbers speak for themselves. Opportunities to do meaningful biomedical, translational or clinical research at Duke-NUS have never been better.

Explore research career opportunities at **www.duke-nus.edu.sg/careers** or read about our myriad research programmes at **www.duke-nus.edu.sg/research**





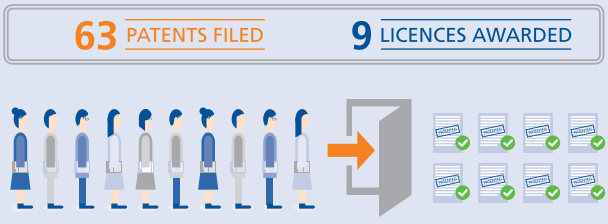
PUBLISHED OVER **2,800** PAPERS IN INTERNATIONAL PEER-REVIEWED JOURNALS



>S$350 MILLION GARNERED IN RESEARCH FUNDING LOCALLY AND OVERSEAS



>95 SYNERGISTIC RESEARCH ALLIANCES



RESEARCH AWARDS WON

11 MOH NMRC SINGAPORE TRANSLATIONAL RESEARCH (STaR) INVESTIGATOR AWARDS*

25 MOH NMRC CLINICIAN SCIENTIST AWARDS*

24 MOH NMRC TRANSITION AWARDS*

*Duke-NUS faculty and faculty holding primary academic appointment at Duke-NUS

All information accurate as at **1 July 2016**

DUKE-NUS MEDICAL SCHOOL
8 College Road, Singapore 169857 **Email:** hr@duke-nus.edu.sg **Website:** www.duke-nus.edu.sg

Isotopic composition of plant water sources

ARISING FROM J. Evaristo, S. Jasechko & J.J. McDonnell *Nature* **525,** 91–94 (2015); doi:10.1038/nature14983

In their study, Evaristo *et al.*[1] collected an extensive data set on the basis of which they statistically determined the isotopic compositions of the plant water source ($\delta^{18}O_{intersect}$ and $\delta^{2}H_{intersect}$, called respectively $\delta^{18}O_{intercept}$ and $\delta^{2}H_{intercept}$ in their paper) as the *x* and *y* coordinates in ($\delta^{18}O$, $\delta^{2}H$) space of the intersection between the local meteoric water line (LMWL) and the plant xylem water 'evaporation line' (EL) for a range of climates and vegetation types. Evaristo *et al.*[1] showed that for 80% of their referenced sampling sites, the mean value of the groundwater hydrogen isotopic composition ($\delta^{2}H_{GW}$) was statistically different from $\delta^{2}H_{intersect}$, supporting their hypothesis that the precipitation sources for groundwater recharge are different from the precipitation sources for plant water uptake. However, we consider that Evaristo *et al.*[1] made a mistake in their equation (2), such that their analysis of rainfall segregation between plant transpiration and groundwater is not valid. This result does however not bring into question the paper's conclusion[1] of hydrological separation formulated on the basis of the precipitation offset. There is a Reply to this Brief Communication Arising by Evaristo, J., Jasechko, S. & McDonnell, J.J. *Nature* **536,** http://dx.doi.org/10.1038/nature18947 (2016).

The equation used in ref. 1 for the LMWL at a given sampling site is:

$$\delta^{2}H = a\delta^{18}O + b$$

with *a* and *b* the slope and intercept with the *y* axis of the LMWL, respectively. Similarly for the xylem water EL:

$$\delta^{2}H = m\delta^{18}O + n$$

with *m* and *n* the slope and intercept with the *y* axis of the xylem water EL, respectively. The intercept of the xylem water EL with the *y* axis is thus:

$$n = \delta^{2}H - m\delta^{18}O$$

The *x* and *y* coordinates of the intersection between these two lines (respectively $\delta^{18}O_{intersect}$ and $\delta^{2}H_{intersect}$) are obtained in a straightforward manner (equations (1) and (2) below):

$$a\delta^{18}O_{intersect} + b = m\delta^{18}O_{intersect} + n$$

$$\delta^{18}O_{intersect} = \frac{n-b}{a-m} \quad (1)$$

$$\delta^{2}H_{intersect} = a\delta^{18}O_{intersect} + b \quad (2)$$

leading to:

$$\delta^{18}O_{intersect} = \frac{\delta^{2}H_{intersect} - b}{a} \quad (3)$$

Equation (3) above corresponds to equation (3) in ref. 1. Yet, it can be demonstrated that their equation (2), that is,

$$\delta^{2}H_{intersect,Evaristo} = \delta^{2}H - m\delta^{18}O$$

does not give the right $\delta^{2}H_{intersect}$, as $\delta^{18}O_{intersect}$ is missing and should therefore be rewritten as:

$$\delta^{2}H_{intersect} = \delta^{2}H - m(\delta^{18}O - \delta^{18}O_{intersect})$$

We applied equations (1) and (2) above to the complete data set provided by the authors of ref. 1: we used their methodology together with our set of corrected equations to identify for each site the slope of the EL and the coordinates of the intersection between the EL and the LMWL ($\delta^{18}O_{intersect}$ and $\delta^{2}H_{intersect}$).

Out of the 47 available sites, $\delta^{18}O_{intersect}$ and $\delta^{2}H_{intersect}$ could be computed for 46 sites. One site (ID 28) had only one plant xylem water

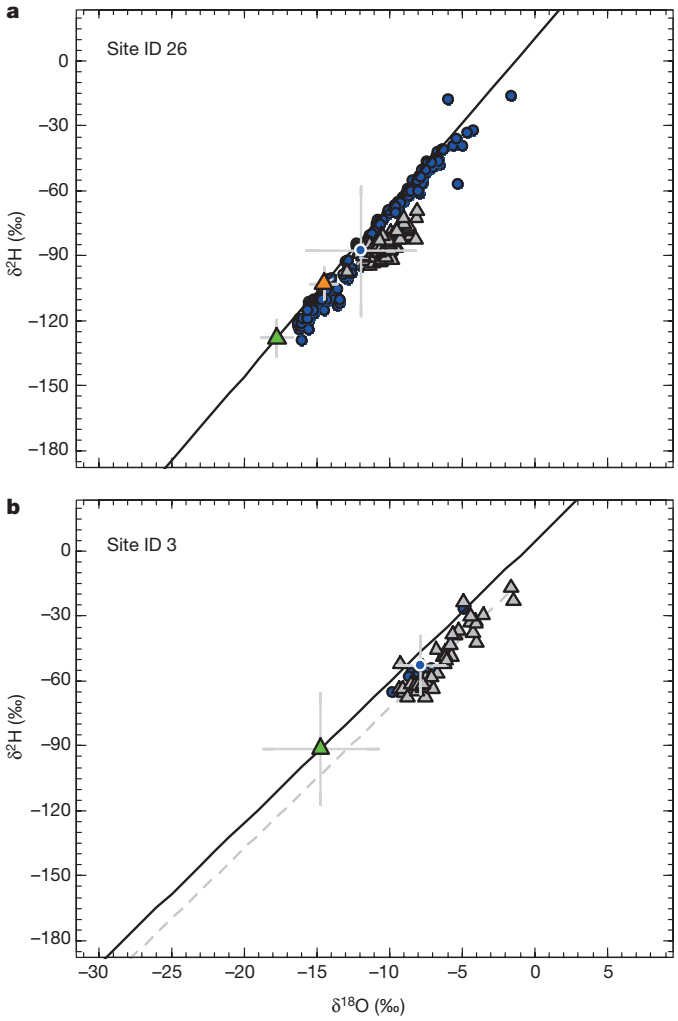


Figure 1 | Comparison between the erroneous and correct isotopic compositions of plant water source precipitation for two sites. The erroneous source precipitation oxygen and hydrogen isotopic compositions (green triangle in ($\delta^{18}O$, $\delta^{2}H$) space) were calculated with equations (2) and (3) of Evaristo *et al.*[1], whereas the correct ones (orange triangle) were calculated with equations (1) and (2) from the present work. The correct orange triangle is located at the intersection between the LMWL (solid black line) and the EL (dashed grey line). Shown are values for plant xylem water (grey triangles), mean groundwater (blue circle with error bars) and groundwater (blue circles) isotopic compositions plotted in ($\delta^{18}O$, $\delta^{2}H$) space. **a**, An example from Oregon, USA (same example as in extended data figure 2 of Evaristo *et al.*[1], site ID 26). **b**, An example of an inconsistent data set, from Arizona, USA (site ID 3). Error bars, ± 1 s.d.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Table 1 | Site-specific plant water source and groundwater isotopic compositions

Site ID	Plant water source $\delta^2H_{intersect}$ (‰)	Plant water source $\delta^{18}O_{intersect}$ (‰)	Groundwater δ^2H_{GW} (‰)	Groundwater $\delta^{18}O_{GW}$ (‰)	Criterion 1 (1 = yes, 0 = no)	Criterion 2 (1 = yes, 0 = no)	Plant versus GW
1	−39 (10.8)	−6.9 (1.5)	−25.5 (1.8)	−4.4 (0.3)	1	1	**
2	−105.3 (7.7)	−13.7 (1.1)	−93.8 (21.8)	−12.1 (3.4)	1	1	**
3	−2,653.4 (1416)	−407.7 (217.2)	−52.6 (3.1)	−7.8 (1.3)	1	0	NA
4	−137 (23.9)	−18.4 (3)	−80.7 (0.2)	−11.1 (0)	1	1	**
5	−93.8 (8.2)	−13.9 (1.3)	−81.5 (8.6)	−10.7 (0.9)	1	1	NS
6	−92.5 (6.9)	−12.2 (0.9)	−49.8 (13)	−7.4 (1.8)	1	1	**
7	−233.7 (21)	−31.7 (2.8)	−74.2 (7.6)	−9.9 (1.1)	1	1	NA
8	−399.7 (128.9)	−57.7 (18.1)	−43 (5.1)	−6.4 (0.5)	1	1	NA
9	−57.1 (18.9)	−7.7 (2.5)	−28.8 (19)	−4.8 (1.9)	1	1	**
10	−41.5 (0)	−6.6 (0)	−29.2 (2.1)	−4.3 (1.2)	1	1	**
11	−28.8 (7.2)	−5.1 (1.1)	−24.7 (0.4)	−4.7 (0.4)	1	1	**
12	−24.1 (3.3)	−5.4 (0.6)	−21.2 (16)	−4.2 (2)	1	1	NS
13	−72.5 (13)	−9.5 (1.7)	−67 (11.9)	−9.6 (1.6)	1	1	NS
14	−72.5 (107)	−9.8 (12.9)	−65.2 (6.2)	−9.2 (1.1)	1	1	NS
15	−64.9 (12.8)	−11.3 (2.1)	−40.1 (2.6)	−7 (0.5)	1	1	**
16	−83.7 (4.8)	−12.8 (0.7)	−67 (22.2)	−8.8 (2.8)	1	1	**
17	11.6 (37.9)	0.2 (4.7)	−46.6 (5.5)	−6.7 (0.4)	0	1	NA
18	−100.4 (39.7)	−14.3 (5.6)	−53.4 (9.2)	−8.3 (1.5)	1	1	**
19	−77.2 (11.9)	−11.1 (1.6)	−62.8 (4.8)	−8.3 (0.8)	1	1	**
20	−17.5 (14.3)	−1.6 (2.2)	−74.7 (6.5)	−9.7 (0.5)	0	1	NA
21	−415.2 (183.1)	−81.6 (35.7)	−29.3 (2.8)	−4 (1.1)	1	0	NA
22	−93.2 (0)	−13.6 (0)	−85.1 (5.9)	−12.7 (1)	1	1	**
23	−551.2 (90.3)	−69.9 (11.2)	−106 (16.8)	−14.1 (2)	1	0	NA
24	−893.1 (469.6)	−121.3 (63.3)	−89.9 (12.6)	−12.1 (1.6)	1	0	NA
25	−97.3 (40.6)	−12.8 (5.2)	−104.7 (5.4)	−14.3 (0.8)	1	1	NS
26	−103.8 (11.4)	−14.6 (1.5)	−87.8 (53)	−11.9 (6.5)	1	1	NS
27	−89.6 (4.3)	−11.9 (0.5)	−85.1 (5.9)	−12.7 (1)	1	1	NS
28	NA (NA)	NA (NA)	−56.1 (1.1)	−8.8 (0.1)	NA	NA	NA
29	−128.8 (15.2)	−16.5 (2.2)	−108.3 (20.8)	−13.6 (3.6)	1	1	**
30	−81.1 (22.3)	−10.6 (2.7)	−91.7 (6.5)	−11.5 (0.5)	1	1	**
31	−38.7 (11)	−6.1 (1.4)	−35 (0.2)	−4.9 (0)	1	1	NS
32	−586.4 (0)	−83.9 (0)	−28.4 (2.5)	−5 (0.3)	1	0	NA
33	−69.9 (34.7)	−9.8 (4.6)	−73.6 (10.9)	−10 (1)	1	1	NS
34	−49.4 (7.2)	−7.6 (0.9)	−34.6 (5.2)	−5.2 (1.3)	1	1	**
35	−49.9 (10.5)	−8.4 (1.6)	−17 (1.3)	−4.5 (0.3)	1	1	**
36	−9.7 (1)	−2.2 (0.2)	−26.4 (2.9)	−4.5 (0.2)	0	1	NA
37	−43.7 (9.2)	−6.8 (1.2)	−29.3 (2.8)	−4 (1.1)	1	1	**
38	−123.5 (36.9)	−17.4 (4.9)	−62.8 (15.8)	−9 (2.1)	1	1	**
39	−25.2 (0.8)	−3.9 (0.1)	−38.5 (15.9)	−5.7 (2)	1	1	**
40	−72 (22.7)	−12.2 (3.5)	−21.6 (14)	−4.5 (0.6)	1	1	**
41	−168.1 (49.6)	−24.2 (6.5)	−14.4 (2.7)	−3 (0.3)	1	1	**
42	−118.7 (22.5)	−15.7 (2.6)	−48.5 (7.8)	−7.6 (0.8)	1	1	**
43	−51.1 (16.4)	−7.8 (1.9)	−9.9 (7.5)	−2.6 (0.8)	1	1	**
44	−359.1 (145.8)	−46 (17.8)	−62.8 (15.8)	−9 (2.1)	1	1	NA
45	−97.2 (20.1)	−12.4 (2.2)	−46.6 (5.5)	−6.7 (0.4)	1	1	**
46	−159.2 (129.2)	−21.1 (15.9)	−14.4 (2.7)	−3 (0.3)	1	1	**
47	−50.3 (8.5)	−7.5 (1.1)	−14.4 (2.7)	−3 (0.3)	1	1	**

This table shows the hydrogen and oxygen isotopic compositions of the plant water source ($\delta^2H_{intersect}$ and $\delta^{18}O_{intersect}$) for each site, presented as median (interquartile range), calculated using equations (1) and (2) of the present work. Values of groundwater hydrogen and oxygen isotopic compositions (δ^2H_{GW} and $\delta^{18}O_{GW}$) are also presented as median (interquartile range). Columns 6 and 7 are criteria for the calculated intersection value: criterion 1, $\delta^2H_{intersect} < \max(\delta^2H_{plant})$; and criterion 2, $\delta^2H_{intersect} > -200‰$. NA, not applicable, that is, when at least one criterion was not met by site data, or when data were lacking to perform the statistical analysis. **Denotes statistically significant difference between plant water source and groundwater (GW) hydrogen stable isotopic compositions using the non-parametric Wilcoxon rank sum test ($\alpha = 0.05$). NS, not significant.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

isotopic measurement and could therefore not be used for regression analysis.

Figure 1a shows one example of the difference between the original estimate of Evaristo *et al.*[1] (green triangle) and our revised estimate (orange triangle) of the plant water source (site ID 26). For most of the sites, our estimates differ considerably from those of Evaristo *et al.*[1] (see Extended Data Figs 1, 2 and 3).

Extended Data Figs 1, 2 and 3 also reveal that the data set is extremely heterogeneous in terms of the number of sampled points for plant and groundwater, with sometimes inconsistent data leading to unrealistic values of $\delta^{18}O_{intersect}$ and $\delta^{2}H_{intersect}$ for certain sites. Therefore, we applied two non-exclusive criteria to assess the consistency of the calculated intersection values: criterion 1, $\delta^{2}H_{intersect} < \max(\delta^{2}H_{plant})$; and criterion 2, $\delta^{2}H_{intersect} > -200‰$. The first criterion implies that $m < a$, while the second criterion evaluates whether the plant water source hydrogen isotopic composition value is realistic. Eleven sites failed at least one of the two criteria (IDs 3, 7, 8, 17, 20, 21, 23, 24, 32, 36 and 44). Figure 1b shows one example of an inconsistent data set (site ID 3).

Results of this analysis (summarized in Table 1) show that at 26 sites, where data were consistent, $\delta^{2}H_{GW}$ was statistically different from $\delta^{2}H_{intersect}$ using the non-parametric Wilcoxon rank sum test ($\alpha = 0.05$). In conclusion, rainfall segregation (as defined by Evaristo *et al.*[1]) could be observed for only 57% of the sites of the authors' data set and at 74% of the sites with consistent estimates of the intercept as defined by our two criteria.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Mathieu Javaux[1,2,3], Youri Rothfuss[1], Jan Vanderborght[1], Harry Vereecken[1] & Nicolas Brüggemann[1]
[1]Institute of Bio- and Geosciences, IBG-3 Agrosphere, Forschungszentrum Jülich GmbH, Jülich 52425, Germany.
email: m.javaux@fz-juelich.de
[2]Earth and Life Institute, Université catholique de Louvain (UCL), Louvain-la-Neuve 1348, Belgium.
[3]Department of Land, Air and Water Resources, University of California Davis, Davis, California 95616, USA.

Received 25 January; accepted 6 June 2016.

1. Evaristo, J., Jasechko, S. & McDonnell, J. Global separation of plant transpiration from groundwater and streamflow. *Nature* **525,** 91–94 (2015).

Author Contributions M.J. found the mistake in the original paper. Y.R. performed the statistical analyses. M.J., Y.R., J.V., H.V. and N.B. discussed the results and wrote the manuscript.

Competing Financial Interests Declared none.

doi:10.1038/nature18946

Evaristo *et al.* reply

REPLYING TO M. Javaux, Y. Rothfuss, J. Vanderborght, H. Vereecken & N. Brüggemann *Nature* **536,** http://dx.doi.org/10.1038/18946 (2016)

In the accompanying Comment[1], Javaux *et al.* correct a mistake in equation (2) of our work[2]; as they point out[1], the mistake does not impact the central conclusion of our paper that ecohydrological separation is widespread[2]. However, equations (2) and (3) in ref. 2 calculate the source precipitation value of xylem water as the point where the xylem water evaporation line (EL) intersects the local meteoric water line (LMWL). In so doing, Javaux *et al.*[1] note that the mistake affects our finding[2] that "at 80% of the sites, the precipitation that supplies groundwater recharge and streamflow is different from the water that supplies parts of soil water recharge and plant transpiration".

There are two key points in our response.

(1) We recognize the mistake now noted in equation (2) and thank Javaux *et al.*[1] for this correction. These authors[1] find that rainfall segregation could be observed at only 74% of the sites (as defined by the two criteria in ref. 1), and not 80% as we originally reported[2].

(2) Our work[2] presented evidence for ecohydrological separation based on a meta-analysis of isotopic dual liquid water isotope data ($\delta^{2}H$ and $\delta^{18}O$) from 47 studies. This conclusion is supported by studies that analysed water vapour isotope data from the Tropospheric Emissions Spectrometer aboard NASA's Aura satellite[3] and by global differences between annual precipitation and groundwater isotope compositions[4,5]. These global-in-scale lines of evidence support earlier field evidence[6,7] that ecohydrological separation (defined as plants using water of a character different to that of mobile water found in soils, groundwater and streamflow) is widespread, and is the rule rather than the exception. Ecohydrological separation was calculated using equation (1) in ref. 2.

It must be understood that equation (1) in ref. 2 is independent of the source precipitation analysis, which was calculated using equations (2) and (3) in that paper. Therefore, any issue with equation (2) in our paper, like the one raised by ref. 1, does not affect the ecohydrological separation conclusion.

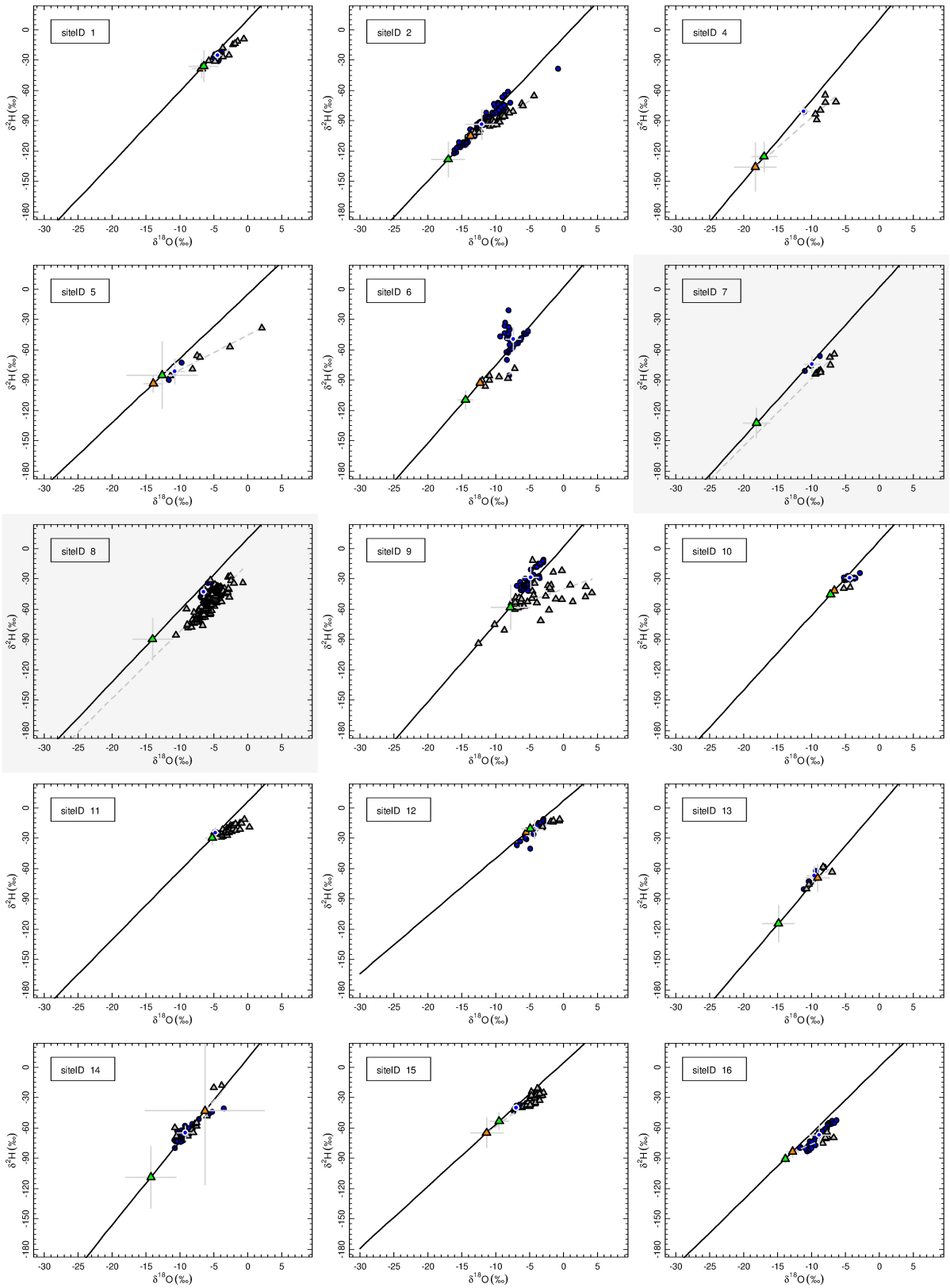
We hope that this exchange will generate further interest in the use of stable O and H isotopes in plant water relation studies.

J. Evaristo[1], S. Jasechko[2] & J. J. McDonnell[1,3,4]
[1]Global Institute for Water Security and School of Environment and Sustainability, University of Saskatchewan, Saskatoon, Saskatchewan S7N 3H5, Canada.
email: jaivime.evaristo@usask.ca
[2]Department of Geography, University of Calgary, Calgary, Alberta T2N IN4, Canada.
[3]School of Geosciences, University of Aberdeen, Aberdeen AB34 3FX, UK.
[4]Department of Forest Engineering, Resources and Management, Oregon State University, Corvallis, Oregon 97331, USA.

1. Javaux, M., Rothfuss, Y., Vanderborght, J., Vereecken, H. & Brüggemann, N. Isotopic composition of plant water sources. *Nature* **536**, http://dx.doi.org/10.1038/nature18946 (2016).
2. Evaristo, J., Jasechko, S. & McDonnell, J. J. Global separation of plant transpiration from groundwater and streamflow. *Nature* **525,** 91–94 (2015).
3. Good, S. P., Noone, D. & Bowen, G. Hydrologic connectivity constrains partitioning of global terrestrial water fluxes. *Science* **349,** 175–177 (2015).
4. Jasechko, S. & Taylor, R. G. Intensive rainfall recharges tropical groundwaters. *Environ. Res. Lett.* **10,** 124015 (2015).
5. Jasechko, S. *et al.* The pronounced seasonality of global groundwater recharge. *Wat. Resour. Res.* **50,** 8845–8867 (2014).
6. Brooks, J. R., Barnard, H. R., Coulombe, R. & McDonnell, J. J. Ecohydrologic separation of water between trees and streams in a Mediterranean climate. *Nat. Geosci.* **3,** 100–104 (2010).
7. Evaristo, J., McDonnell, J. J., Scholl, M. A., Bruijnzeel, L. A. & Chun, K. P. Insights into plant water uptake from xylem-water isotope measurements in two tropical catchments with contrasting moisture conditions. *Hydrol. Process.* http://dx.doi.org/10.1002/hyp.10841 (in the press).

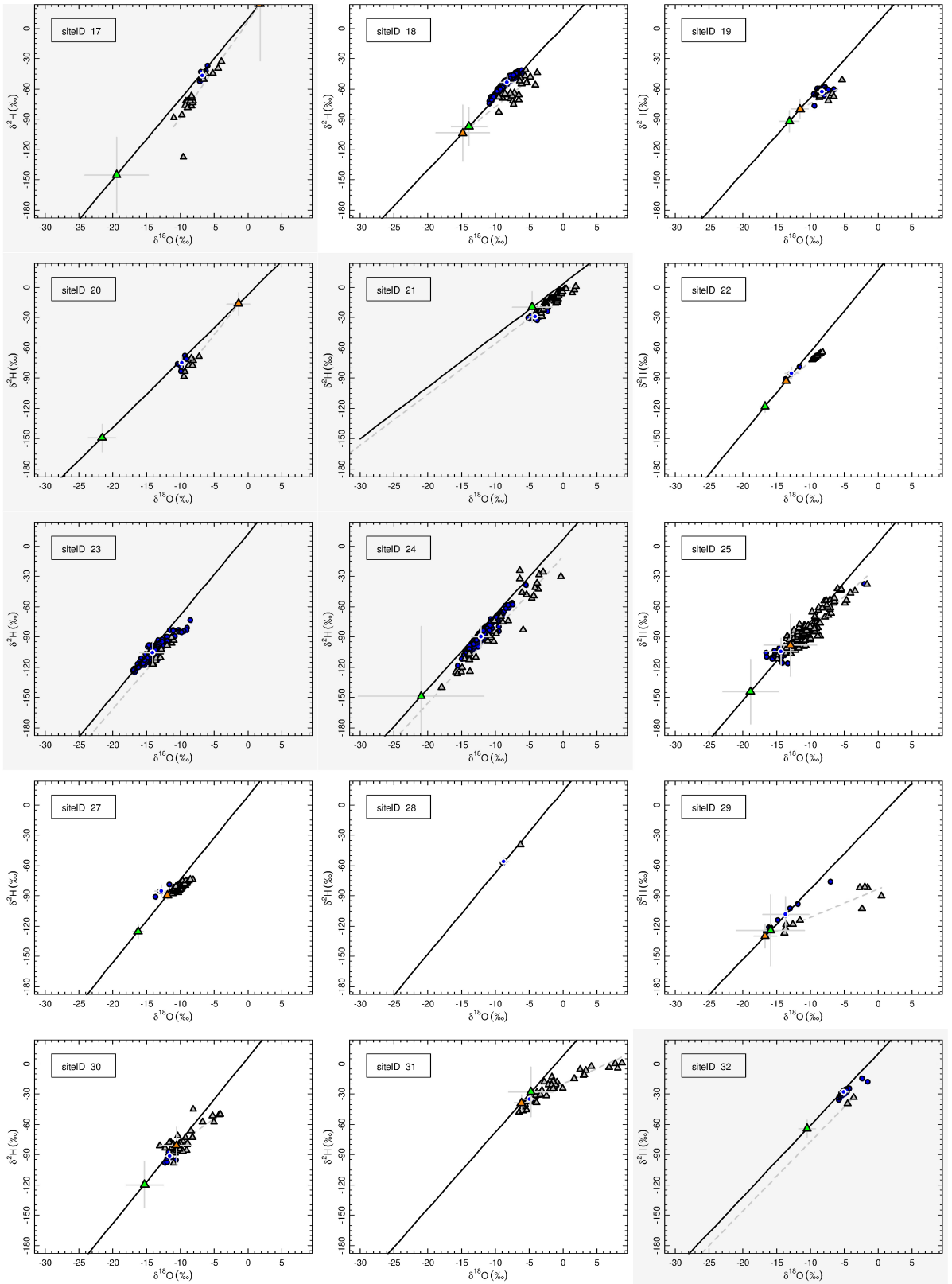
doi:10.1038/nature18947

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



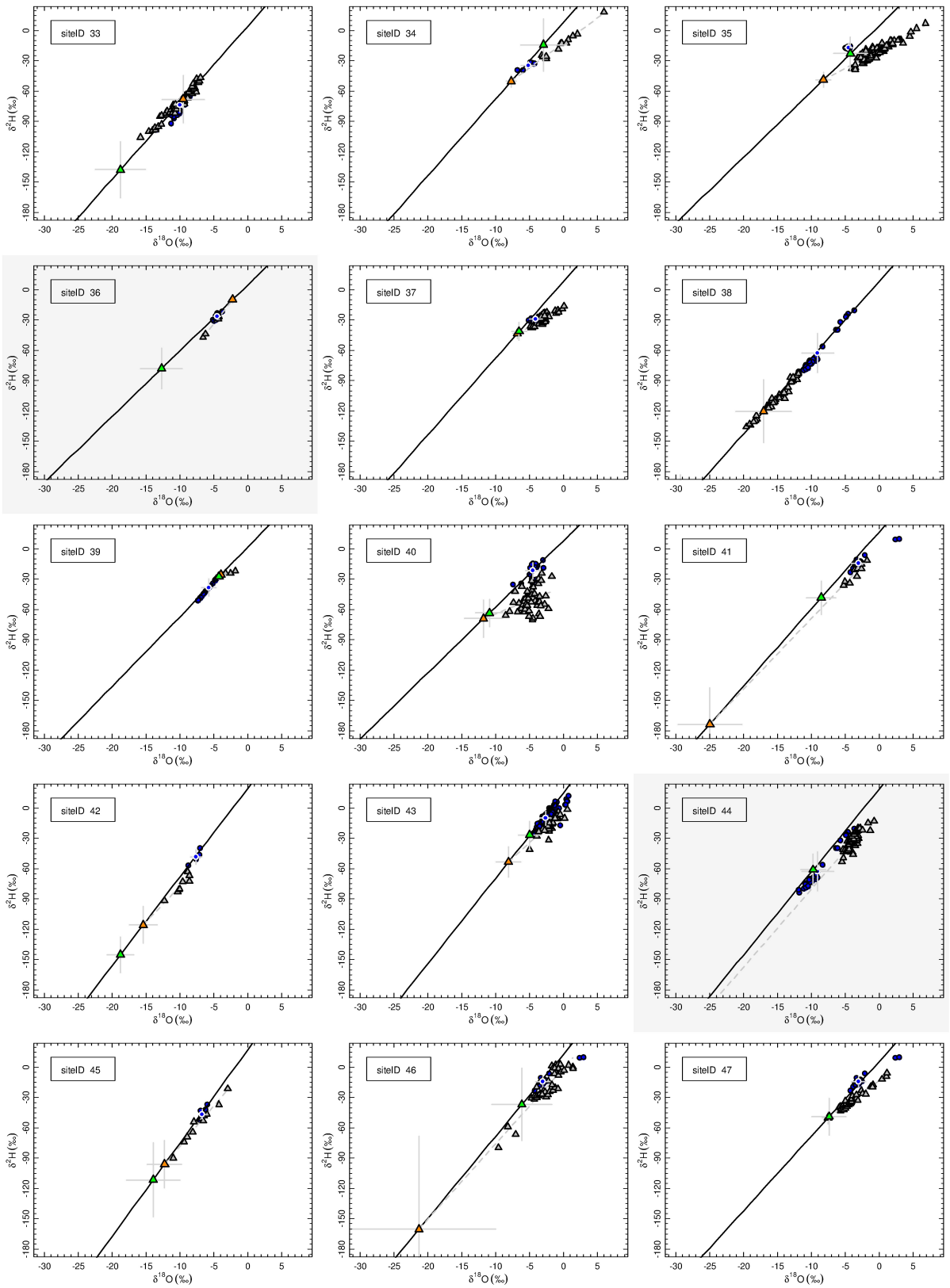
Extended Data Figure 1 | Comparison of plant water source isotopic compositions according to corrected and original formulations across sites ID 1 to ID 16. Comparison of plant water source isotopic compositions at sites ID 1 to 16 (site ID 3 is in Fig. 1b of the present paper) according to the corrected (equations (1) and (2) of the present work, orange triangle) and the original (equations (2) and (3) of Evaristo *et al.*[1], green triangle) formulation (error bars, ± 1 s.d.). The correct orange triangle is located at the intersection between the LMWL (solid black line) and the EL (dashed grey line). Grey plots are sites where at least one of the two consistency criteria was not met.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 2 | As Extended Data Fig. 1 but for sites ID 17 to 32. Site ID 26 is in Fig. 1a of the present paper.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 3 | As Extended Data Fig. 1 but for sites ID 33 to 47.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

AT MONASH, EXCELLENCE AND EQUITY IS NOT WHAT WE DO, IT IS WHO WE ARE.



At Monash, we are committed to ensuring our staff and students reflect the world we are working towards: diverse, inclusive, innovative and sustainable. We actively support women in senior roles, especially in the fields of science, technology, engineering, maths and medicine (STEMM). Our commitment is demonstrated by our participation in the Science in Australia Gender Equity pilot of Athena SWAN.

Above all, we value excellence, no matter where you're from. As the largest university in Australia, with an international reputation for driving change, you can be assured that we have a wealth of resources to support you and your research. Join us at Monash University, where great minds come together to shape the future.

WGEA Employer of Choice for Gender Equality

Athena SWAN Member

monash.edu/jobs

MONASH University

CRICOS provider: Monash University 00008C.







Ningbo University

Seeking bright minds

Located in the historical port city of Ningbo in eastern China, Ningbo University is a burgeoning comprehensive university co-established by the Chinese Ministry of Education, Zhejiang provincial government and Ningbo municipal government. Young and dynamic, Ningbo University is already ranked among the top 100 universities in China.

Ningbo University is actively seeking talented researchers to strengthen its faculty team.

Openings for academic leaders

Requirements:

- A doctoral degree from an overseas institution is expected, along with at least three years of work experience conducting research overseas; for those who have obtained their doctoral degree from a domestic institution, at least three years of overseas teaching or research experience is a must
- Experience working as a tenured professor or equivalent in a well-known university or research institution overseas (associate professor experience is fine for young candidates from top universities or institutions); generally, candidates should qualify for the national Thousand Talents Program
- A proven track record of achievements in a specialized research field, with the potential to become an academic or technical leader in the field
- Ability to work full-time on site, and preferably under 50 years old

Openings for top young scientists

Requirements:

- A doctoral degree from an overseas institution is preferred, along with at least three years of post-graduate research experience overseas; those with doctoral degrees from domestic institutions must have at least three years of experience conducting research or teaching overseas
- Experience working full-time in a well-known university or research institution overseas, conducting research or teaching; generally, candidates should qualify for the national Thousand Young Talents Program or the provincial Thousand Talents Program
- Ability to work full-time on site, and preferably under 45 years old

Openings for excellent doctoral researchers

Requirements:

- A doctoral degree from an overseas institution is preferred, along with at least three years of work experience conducting research overseas; those with doctoral degrees from domestic institutions should have at least three years of experience conducting research or teaching overseas
- A track record of publication experience, with at least one paper published in Social Sciences Citation Index or Arts & Humanities Citation Index journals for candidates in humanities and social sciences fields; two or more papers published in Science Citation Index-listed journals or at least one publication in a top journal for candidates in natural sciences fields
- Ability to work full-time at the university

Compensation

Generous compensation packages will be available. For excellent doctoral researchers, the successful candidate will receive a settling-in allowance of 600,000 (180,000+420,000) RMB. Those with four or more publications in top journals are eligible to be hired as associate professors, and will receive a settling-in allowance of 800,000 (600,000+200,000) RMB.

Application procedure

Please submit a completed application form, a curriculum vitae and a cover letter, along with other relevant supporting materials via e-mail to: rsc@nbu.edu.cn.

For additional information regarding the application, such as the number of openings, please visit: http://www.nbu.edu.cn/shizi.

宁波大学
NINGBO UNIVERSITY

Contact

E-mail: rsc@nbu.edu.cn
Tel. no: +86-574-87600288
Website: www.nbu.edu.cn/english

Advertiser retains sole responsibility for content

NANYANG TECHNOLOGICAL UNIVERSITY



Here are 3 good reasons to join NTU Singapore,
THE WORLD'S TOP YOUNG UNIVERSITY

THE WORLD'S BEST YOUNG UNIVERSITY IS LOOKING FOR THE WORLD'S MOST PROMISING PROFESSORS

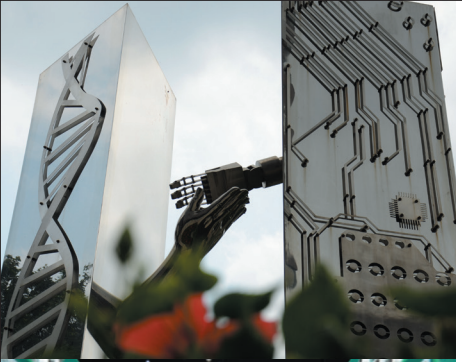
Pursue your highest aspirations at the world's fastest-rising young university that is also in the top 13 of the global university league

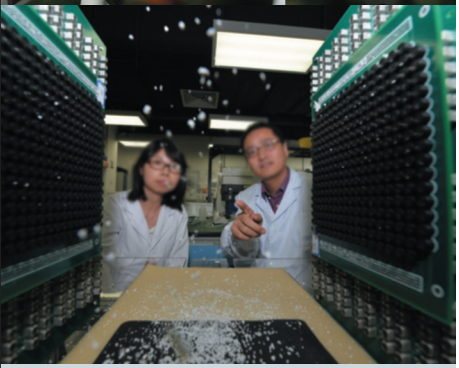
Reap results at this research-intensive university that leads the top Asian universities in normalised research citation impact (Thomson Reuters InCites 2015) and is ranked 32nd globally among universities in Nature Index 2016

Rise in your chosen field as an elite **Nanyang Assistant Professor** with a start-up research grant of up to US$727,000 (S$1 million), an attractive remuneration package and a tenure-track appointment

Apply for the 2017 Nanyang Assistant Professorship
If you are an early-career researcher (postdoctoral fellow or equivalent), and are ready to lead your research group independently, write to us at nanyangprofessorship@ntu.edu.sg or visit www.ntu.edu.sg/nap to apply.
Application deadline: 9 October 2016

www.ntu.edu.sg





Job openings at SIAT

The Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences

The Shenzhen Institutes of Advanced Technology (SIAT), Chinese Academy of Sciences (CAS), was jointly established by CAS, the Chinese University of Hong Kong and the Shenzhen Municipal Government in 2006. SIAT has earned an outstanding reputation for translating basic research into industrial applications, thanks to its interdisciplinary teams of scientists and engineers. At the interface between biological technologies and information technologies, SIAT is well positioned for innovation in the twenty-first century.

SIAT is hiring for the following positions:

- Researchers for the National Thousand Talents Program
- Researchers for the CAS Hundred Talents Program
- Distinguished research fellows
- Professors, associate professors, and assistant professors

Research fields include:

- Advanced biomedical imaging and therapeutics technologies
- Biomedical devices and medical robotics
- Neural engineering and rehabilitation
- Biomaterials and tissue engineering
- Biomedical informatics and genome sequencing
- Micro- nano- and bio-engineering
- Nanomedicine, antibodies and immunology
- Robotics and intelligent systems
- High-performance and cloud computing
- Information and communications technologies
- Data mining
- Precision engineering
- Water-treatment technologies
- Neurobiological mechanisms of brain disorders

Interested candidates should contact:

Rui Wang
Vice director of the Department of Human Resources and Education
Tel: +86-755-86392095
E-mail: rui.wang@siat.ac.cn

中国科学院深圳先进技术研究院
SHENZHEN INSTITUTE OF ADVANCED TECHNOLOGY
CHINESE ACADEMY OF SCIENCES

Southwest Jiaotong University

Invitation to apply for academic positions

Located in Chengdu, the capital of China's Sichuan province, Southwest Jiaotong University (SWJTU) is an elite comprehensive university founded in 1896. Excelling in engineering, SWJTU also covers the sciences and arts with its 19 faculties, institutes and centres. Its 2,600 faculty and staff offer complete undergraduate, graduate and postdoctoral programmes.

SWJTU cordially invites researchers to apply for positions across all its disciplines.

Positions and requirements

High-level Talented Leaders

Applicants should be qualified for national top talents programmes such as the Program of Global Experts, Cheung Kong Scholars and National Science Fund for Distinguished Young Scholars.

Young Leading Scholars

Applicants should preferably be qualified for the following programmes:

- National Thousand Young Talents Program
- Top Young Talents of National Special Support Program
- National Natural Science Foundation of China Excellent Young Scholars Program

Applicants should also have a strong team spirit and leadership qualities, outstanding academic achievements, a broad academic vision and experience with international cooperation along with the potential to become leading academic researchers.

Excellent Young Academic Backbones

Applicants should preferably be under 40 years old with degrees from first-class universities or institutes in China or overseas.

Excellent Postdoctoral Fellows

Applicants should preferably be under 35 years old and have the potential to become excellent academic researchers.

Compensation

Salaries will be highly competitive, commensurate with qualifications and experience. Other benefits include settling-in allowances, housing subsidies, child education support, start-up funds, assistance in establishing scientific platforms and research groups, along with international training and promotion opportunities. Detailed packages are negotiable.

Application procedure

Please send a curriculum vitae, copies of academic credentials, a publication list with abstracts of selected published papers, a research plan, a teaching statement and the contact details of three referees to the Human Resources Department.



For inquiries, please contact Mr. Yu Wang, Ms. Ye Zeng or Ms. Qingya Wang

Telephone: +86-28-66366202
E-mail: talent@swjtu.edu.cn
Address: Human Resources Department, SWJTU, Western Park of High-Tech Zone, Chengdu, Sichuan, China, 611756
Website: www.swjtu.edu.cn



Seeking Talented Researchers Worldwide

Located in the historical and picturesque city of Hangzhou, Zhejiang University is a prestigious higher education institution in China. With a long history that can be traced back to the Qiushi Academy founded in 1897, the university enjoys a good reputation at home and abroad. After more than 100 years of development, it has become a comprehensive research university with unique characteristics. It has expanded to include 7 campuses and 36 colleges and schools, and its research and training programmes span 12 academic disciplines. Of the 22 disciplines listed in Thomson Reuters' Essential Science Indicators (ESI) database, Zhejiang University has 18 ranked in the top 1% in the world.

Zhejiang University has always stayed true to its commitment to cultivating talent with excellence, advancing science and technology, serving the well-being of society, and promoting advanced culture. This spirit is best captured by the university's motto of "seeking the truth and pioneering new trails".

Following the educational philosophy of putting people first, cultivating all-round competence in students, seeking the truth and pioneering new trails in search of excellence, Zhejiang University is committed to cultivating future leaders who have international outlooks. It is also leading the country in educational reforms. A large number of outstanding figures in modern Chinese history have studied or conducted research at the university. More than 160 of its alumni have gone on to become members of the Chinese Academy of Sciences or the Chinese Academy of Engineering. With a rich campus culture, advanced teaching facilities and a wide range of international exchange opportunities, the university is keen to create conditions to facilitate students' development.

Aiming to become a world-class university, Zhejiang University also focuses on research innovation. It has launched a number of international academic platforms and has attracted many talented researchers in various disciplines. To further strengthen its research teams, the university has recently launched the Hundred Talents Program, which seeks to encourage bright minds to join Zhejiang University from home and abroad.

Advertiser retains sole responsibility for content

MULTIPLE OPENINGS FOR THE HUNDRED TALENTS PROGRAM

The Hundred Talents Program is a highly competitive, well-funded programme that covers all colleges and departments at Zhejiang University. Successful candidates are expected to establish internationally competitive and independent research programmes, focusing on cutting-edge science in their respective fields. They are required to work full-time and deliver excellent research and teaching at Zhejiang University, following the spirit of "seeking the truth and pioneering new trails".

Required qualifications

Candidates should have:

- A demonstrated commitment to excellence in teaching and research;
- A proven track record of academic achievement, comparable to those of assistant professors or associate professors in world-renowned universities; and
- A doctoral degree from a world-renowned university, preferably with postdoctoral experience. Young and mid-career researchers are particularly encouraged to apply.

What we offer

The successful candidate will be employed as a Zhejiang University 100 Professor, who is qualified to supervise doctoral students. The university will offer:

- A generous compensation package;
- Subsidized housing;
- Internationally competitive startup packages for research; and
- Large office and laboratory space with state-of-the-art facilities.

How to apply

Qualified applicants are strongly encouraged to submit their applications electronically to tr@zju.edu.cn with the following materials in PDF format:

- A comprehensive CV;
- A statement of research and a teaching plan;
- Certificates of academic degrees;
- Three writing samples of major publications;
- A list of three to five references with detailed contact information.

OTHER ACADEMIC OPENINGS

In addition to the Hundred Talents Program, there are a number of talent recruitment programmes and open faculty positions for talented scholars worldwide, listed below.

Full-time positions:

- Zhejiang University Distinguished Professor of Humanities and Social Science
- The 1000 Talent Plan: Long-term Recruitment Program of Foreign Experts, Long-term Innovative Talents and Young Professionals recruitment programmes
- Cheung Kong Scholar Program: Special-term Professor & Young Scholar programmes
- Zhejiang Provincial Program for High-level Overseas Talents
- Zhejiang University Qiushi Distinguished Scholar
- Zhejiang University Leading Professor of Humanities and Social Science

Part-time positions:

- The 1000 Talent Plan: Short-term Recruitment Program of Foreign Experts and Short-term Innovative Talent programme
- High-end Foreign Experts Recruitment Program
- Zhejiang Provincial Program for High-level Overseas Talents (Short-term Innovative Talent)
- Zhejiang University Chair Professors

For detailed application information on our talent recruitment programmes, please visit our website or scan the QR code below: http://talent.zju.edu.cn/redir.php?catalog_id=20038&object_id=40799

Zhejiang University sincerely invites highly talented scientists and scholars to join its great endeavour. Together, we will create a glorious future for Zhejiang University and its people.

Talents Office,
Human Resources Department,
Zijingang Campus, Zhejiang University
No. 866, Yuhangtang Road, Hangzhou,
Zhejiang, 310058, P. R. China

E-mail: tr@zju.edu.cn
Tel: +86-571-88981390
+86-571-88981345
Fax: +86-571-88981976
Website: http://talent.zju.edu.cn



Advertiser retains sole responsibility for content